International Conference on
Data Mining
(SDM23)

April 27–29, 2023

This document was current as of April 12, 2023. Abstracts appear as submitted.

IP1
Interdisciplinary Modeling to Explore Interventions for Disaster Preparation and Response

Preparing and responding to natural disasters requires integrated models of natural processes and human activities across disparate disciplines to create a wholistic picture of potential effects and possible preventive measures. It typically takes many months and even years create valid end-to-end simulation scenarios as the different models need to be configured in consistent ways so their results can be meaningfully interpreted. Models and scenarios must be designed to inform decision makers about potential interventions for disaster mitigation. In this talk, we will introduce the MINT modeling framework that uses AI for integrated modeling, in particular knowledge technologies, machine learning, and constraint reasoning. MINT represents modeling variables unambiguously in order to facilitate model integration, captures model requirements and constraints in order to validate the use of models, and explicitly connects interventions and decisions to modeling scenarios. We will illustrate the use of MINT to study the impacts of climate extremes such as droughts and floods in food insecurity and migration in Sub-Saharan Africa, and to configure controlled fires for wildfire prevention. We will also motivate the need for future research in causality and multi-disciplinary collaboration.

Yolanda Gil
University of Southern California, U.S.
gil@isi.edu

IP2
AI-Enabled Scientific Revolution

During this talk, we will present the findings of a recent NSF sponsored workshop on AI-Enabled Scientific Revolution. AI has the potential to revolutionize science by enabling new approaches that could lead to significant societal advancements. The workshop convened leading researchers in both AI and scientific fields to identify key challenges and opportunities to accelerate the AI for science revolution. An AI-enabled fourth industrial revolution will be crucial in addressing the biggest global societal challenges, such as feeding the growing population sustainably, improving human health, developing renewable energy sources, mitigating and managing climate change, and creating an equitable society. The AI community must focus on foundational research into common sense learning, machine-understandable knowledge bases, and human-machine intelligence. We would like to enable AI insights from one scientific discipline to generalize to others. AI can help cross-domain collaboration of data collection, automated construction of knowledge bases from data, from predictive tasks to descriptive tasks, even beyond explainability, to get closer to the understandings in scientific discovery. Leadership in AI is also critical for the economic prosperity and wellbeing of our country.

Wei Ding
National Science Foundation, U.S.
weiding@nsf.gov

IP3
Towards Trustworthy and Responsible AI/ML

The lack of trust in data-driven solutions has too often prevented their doption and deployment, particularly in applications that directly impact humans in areas such as healthcare, criminal justice and hiring/firing decisions. I will give an overview of the main components of a trustworthy AI/ML system, provide a glimpse of some promising algorithmic as well as policy based approaches, and highlight key challenges that still need to be resolved. Disclaimer: This abstract has not been generated by ChatGPT or associated technologies. Generative AI has greatly amplified the need for trustworthy AI and is generating a host of new challenges even to basic notions such as authorship and verifiability. I will be largely ignoring these dimensions of Trustworthy AI that is the subject of a whole new talk!

Joydeep Ghosh
UT Austin
jghosh@utexas.edu

IP4
Relational Causal Models

Discovery of causal relationships from observational and experimental data is a central problem with applications across multiple areas of scientific endeavor. There has been considerable progress over the past decades on algorithms for eliciting causal relationships through a set of conditional independence queries from data. Much of this work assumes that the data samples are independent and identically distributed (iid). However, data in many real-world settings violate the iid assumption because they are generated by a system of interacting objects e.g., a collaboration network, social network, or entities connected by relations stored in relational databases. Such data violate the iid assumption. Relational causal models offer a formalism for representing and reasoning about causal relationships in relational data. This talk will motivate and introduce relational causal models, and summarize recent progress on relational causal models (i) characterizing the conditional independence relations that hold in a given relational causal model, (ii) sound and complete learning of the structure of a relational causal model using an independence oracle, (iii) quantifying the strength of conditional dependence and testing conditional independence among relational variables from relational data, and (iv) robust learning of the structure of a relational causal model from relational data.

Vasant Honavar
Pennsylvania State University, U.S
vhonavar@psu.edu

CP1
Influence Without Authority: Maximizing Information Coverage in Hypergraphs

In many social networks, besides peer-to-peer communication, people share information via groups. An interesting problem arises in this scenario: for such networks, which are the best groups to start information diffusion so that the number of eventually informed nodes can be maximized? In this study, we formulate a novel information coverage maximization problem in the context of hypergraphs, wherein nodes are connected by arbitrary-size hyperedges (i.e., groups). In contrast to the existing literature on influence maximization, which aims to find authority nodes with high influence, we are interested in identifying the key groups. To address this problem, we present a new information diffusion model for hypergraphs, namely Hypergraph-Independent-Cascade (HIC). HIC generalizes the popular independent cascade model to hypergraphs to allow capturing group-level information

diffusion. We prove the NP-hardness of the proposed problem under HIC, and the submodular monotone property of the information coverage function. Further, inspired by the Degree Discount algorithm, we derive a new heuristic method named Influence Discount (InfDis). Extensive experiments provide empirical evidence for the effectiveness and efficiency of our approach.

Peiyan Li
LMU Munich
lipeiyan@dbs.ifi.lmu.de

Honglian Wang
KTH Royal Institute of Technology
honglian@kth.se

Kai Li
Huawei Noahs Ark Lab
likai210@huawei.com

Christian Böhm
University of Vienna
christian.boehm@univie.ac.at

CP1

Cadence: Community-Aware Detection of Dynamic Network States

Dynamic interaction data is often aggregated in a sequence of network snapshots before being employed in downstream analysis. The two common ways of defining network snapshots are i) a fixed time interval or ii) fixed number of interactions per snapshot. The choice of aggregation has a significant impact on subsequent analysis, and it is not trivial to select one approach over another for a given dataset. More importantly, assuming snapshot regularity is data-agnostic and may be at odds with the underlying interaction dynamics. To address these challenges, we propose a method for community-aware detection of network states (CADENCE) based on the premise of stable interaction time-frames within network communities. We simultaneously detect network communities and partition the global interaction activity into scale-adaptive snapshots where the level of interaction within communities remains stable. We model a temporal network as a node-node-time tensor and use a structured canonical polyadic decomposition with a piece-wise constant temporal factor to iteratively identify communities and their activity levels. We demonstrate that transitions between network snapshots learned by CADENCE constitute network change points of better quality than those predicted by state-of-the-art network change point detectors. Furthermore, the network structure within individual snapshots reflects ground truth communities better than baselines for adaptive tensor granularity.

Maxwell J. McNeil
State University of New York, Albany, U.S.
mmcneil2@albany.edu

Carolina Mattsson, Frank W. Takes
Leiden University
c.mattsson@liacs.leidenuniv.nl,
f.w.takes@liacs.leidenuniv.nl

Petko Bogdanov
University at Albany
pbogdanov@albany.edu

CP1

A Temporal Graphlet Kernel For Classifying Dissemination in Evolving Networks

We introduce the *temporal graphlet kernel* for classifying dissemination processes in labeled temporal graphs. Such processes can be the spreading of (fake) news, infectious diseases, or computer viruses in dynamic networks. The networks are modeled as labeled temporal graphs, in which the edges exist at specific points in time, and node labels change over time. The classification problem asks to discriminate dissemination processes of different origins or parameters, e.g., diseases with different infection probabilities. Our new kernel represents labeled temporal graphs in the feature space of temporal graphlets, i.e., small subgraphs distinguished by their structure, time-dependent node labels, and chronological order of edges. We introduce variants of our kernel based on classes of graphlets that are efficiently countable. For the case of temporal wedges, we propose a highly efficient approximative kernel with low error in expectation. Our experimental evaluation shows that our kernels are computed faster than state-of-the-art methods and provide higher accuracy in many cases.

Lutz Oettershagen
University of Bonn
lutz.oettershagen@cs.uni-bonn.de

Nils Kriege
University of Vienna
nils.kriege@univie.ac.at

Claude Jordan, Petra Mutzel
University of Bonn
claudejordan@gmail.com, petra.mutzel@cs.uni-bonn.de

CP1

Causal Discovery by Graph Attention Reinforcement Learning

Discovery the causal structure graph among a set of variables is a fundamental but difficult task in many empirical sciences. Reinforcement learning based causal discovery from observed data achieves prominent results. However, previous algorithms lack interpretability and efficiency, and ignore the prior knowledge of causal structure. To solve these problems, we propose GARL that leverages graph attention network to embed the structure information and the prior knowledge, and reinforcement learning to search the variable ordering with the best score. GARL takes the structure information and prior knowledge as the computational skeleton of attention to obtain the embedded representation of variables, and then generates variable orderings through the designed ordering model. In addition, the structure information is used to form the DAG corresponding to the variable ordering, which reduces the computational difficulty and improves the efficiency. GARL generates DAGs in the reinforcement learning framework, and uses the score of DAG as the reward to optimize the network structure to search the DAG with the best score. Experimental results on synthetic and real datasets show that our GARL has obvious advantages in multi-node operation efficiency, and competitive results with competitive baselines.

Dezhi Yang, Guoxian Yu, Jun Wang, Zhongmin Yan
Shandong University

kingjun@sdu.edu.cn, gxyu@sdu.edu.cn,
kingjun@sdu.edu.cn, yzm@sdu.edu.cn

Maozu Guo
School of Electrical and Information Engineering
Beijing University of Civil Engineering and Architecture
guomaozu@bucea.edu.cn

CP1

Heterogeneous Graph Contrastive Learning with Meta-Path Contexts and Weighted Negative Samples

Heterogeneous graph contrastive learning has received wide attention recently. Some existing methods use meta-paths, which are sequences of object types that capture semantic relationships between objects, to construct contrastive views. However, most of them ignore the rich meta-path context information that describes how two objects are connected by meta-paths. On the other hand, they fail to distinguish hard negatives from false negatives, which could adversely affect the model performance. To address the problems, we propose MEOW, a heterogeneous graph contrastive learning model that considers both meta-path contexts and weighted negative samples. Specifically, MEOW constructs a coarse view and a fine-grained view for contrast. The former reflects which objects are connected by meta-paths, while the latter uses meta-path contexts and characterizes the details on how the objects are connected. We take node embeddings in the coarse view as anchors, and construct positive and negative samples from the fine-grained view. Further, to distinguish hard negatives from false negatives, we learn weights of negative samples based on node clustering. We also use prototypical contrastive learning to pull close embeddings of nodes in the same cluster. Finally, we conduct extensive experiments to show the superiority of MEOW against other state-of-the-art methods.

Jianxiang Yu, Xiang Li
East China Normal University
jianxiangyu@stu.ecnu.edu.cn, xiangli@dase.ecnu.edu.cn

CP2

Embedding Transfer with Enhanced Correlation Modeling for Cross-Domain Recommendation

Modern internet platforms usually have different scenarios to provide rich recommendation services to meet the diverse demands of users. Cross-domain recommendation (CDR) and multi-domain recommendation (MDR) methods are widely used in such platforms to leverage rich auxiliary information from multiple domains. However, state-of-the-art CDR and MDR methods usually enforce some correlations between source and target embeddings on each user, ignoring the correlations between users in both domains. To address this problem, we adopt a relaxed contrastive loss, that employs the pairwise similarities in the source domain as relaxed labels, enforcing such inter-sample relations are reserved in a weighted manner in the target domain. The basic assumption behind such a design is that users with similar interests should be with similar interacted items in a recommender system, and this work takes a step further to realize and specify such similarity modeling as collaborative signals encoded in both implicit embedding spaces. We validate the effectiveness of the proposed method on a large-scale public dataset and a real production dataset with over 700 million samples. We further experimentally show that the proposed embedding

transfer method is generic, and can be plugged into any existing deep neural networks, such as BERT4Rec. Currently, the proposed embedding transfer techniques have been successfully deployed in the *Guess You Like* in WeTV for the CDR/MDR task.

Shilei Cao, Yujie Lin, Xianli Zhang, Yufu Chen, Zhen Zhu, Yuxin Chen
Tencent
913388146@qq.com, yujaylin@tencent.com,
ryanxzhang@tencent.com, cheeryschen@tencent.com,
stevezhu@tencent.com, danikachen@tencent.com

Buyue Qian
Beijing Chaoyang Hospital, Capital Medical University
qianbuyue@bjcyh.com

Feng Wang, Zang Li
Tencent
feynmanwang@tencent.com, gavinzli@tencent.com

CP2

UFNRec: Utilizing False Negative Samples for Sequential Recommendation

Sequential recommendation models are primarily optimized to distinguish positive samples from negative ones during training. Thus, negative instances sampled from enormous unlabeled data are essential in learning the evolving user preferences. Except for randomly sampling negatives from a uniformly distributed subset, many delicate methods have been proposed to mine negative samples with high quality. However, due to the inherent randomness of negative sampling, false negatives are inevitably collected in model training. Current strategies mainly focus on removing false negatives, which leads to overlooking potential user interests, lack of recommendation diversity, less model robustness, and suffering from exposure bias. We propose a novel method that can Utilize False Negative samples for sequential Recommendation (UFNRec), which thoroughly explores the leverage of false negatives. We first devise a simple strategy to extract false negatives from true negatives and directly reverse the labels of false negatives. To avoid extra noise from reversed samples, we restrict false negatives in the output space by an EMA operation and a consistency regularization loss. To the best of our knowledge, this is the first work to utilize false negatives instead of simply removing them for sequential recommendation. Both offline and online experiment results demonstrate that UFNRec can effectively draw information from false negatives and improve the performance of SOTA models.

Xiaoyang Liu
OPPO
liuxiaoyang@oppo.com

Chong Liu
Tencent
nickcliu@tencent.com

Pinzheng Wang
Soochow University
pzwang@stu.suda.edu.cn

Rongqin Zheng, Lixin Zhang, Leyu Lin
Tencent
q_fight123@163.com, lixinzhang@tencent.com,
goshawklin@tencent.com

Zhijun Chen, Liangliang Fu
OPPO
justy.chen@oppo.com, fuliangliang@oppo.com

CP2

Harvester: Principled Factorization-Based Temporal Tensor Granularity Estimation

Given a tensor that captures temporal data, such as (user, item, time), the way that we set the granularity of the "time' mode can *make or break* our analysis of the data. If we set the granularity to be extremely fine, we end up with a very sparse and high-rank tensor which is essentially incompatible with what virtually all tensor decomposition models expect. Traditionally, this problem has been *avoided* by setting the granularity of the "time' to a "reasonable' aggregation, an approach which has certainly served tensor analysis of temporal methods well so far. However, such an approach requires *tedious* trial-and-error experimentation across a number of such fixed aggregations, where typically the one that provides the most sensible results is retained, and furthermore it is *arbitrary*, since the optimal aggregation over time need not necessarily be uniform. We introduce HARVESTER, the first principled factorization-based approach which seeks to identify the best temporal granularity of a given tensor. Our proposed method leverages multiple aggregated views of the tensor, and a carefully-designed optimization formulation to solve this problem. We extensively evaluate HARVESTER on synthetic and real data, and demonstrate that it consistently produces tensors of very high quality, compared to the state-of-the-art, across the board for a number of different popular quality measures.

Ravdeep Pasricha
University of California, Riverside
Computer Science and Engineering
rpasr001@ucr.edu

Uday Saini
University of California Riverside
usain001@ucr.edu

Nicholas Sidiropoulos
University of Virginia
nikos@virginia.edu

Fei Fang
Carnegie Mellon University
feifang@cmu.edu

Kevin Chan
DEVCOM Army Research Laboratory
kevin.s.chan.civ@army.mil

Evangelos Papalexakis
University of California, Riverside
epapalex@cs.ucr.edu

CP2

Pluggable Deep Thompson Sampling with Applications to Recommendation

Thompson Sampling (TS) is an effective way to deal with the explorationexploitation dilemma for the multi-armed (contextual) bandit problem. Due to the sophisticated relationship between contexts and rewards in real-world applications, neural networks are often preferable to model this relationship owing to their superior representation capacity. In this paper, we study the problem of combining neural networks with TS in a plug-and-play manner. The basic idea is to maintain a posterior distribution over the reward mean relying on the prediction and the deep representation of the neural network for any given context. Specifically, our proposed algorithm, PlugTS (Pluggable deep Thompson Sampling), introduces no change into the network training process, but only requires one additional sampling stage during serving sampling from a univariate Gaussian distribution (by maintaining a positive definite matrix). Theoretically, we prove that PlugTS achieves an $O(\sqrt{T})$ regret bound, which matches the state-of-the-art neural network-based TS, while PlugTS enjoys much lower computational overhead for each iteration. Experimental results on public datasets among traditional classification and recommendation tasks validate the effectiveness and efficiency of PlugTS.

Lu Wang
JD.com
wangwllu@hotmail.com

Yuhai Song
Tsinghua University
songyuhai.syh@gmail.com

Zhe Wang
JD.com
wangzhe488@jd.com

Haoxiang Wang
Tsinghua University
wanghaox19@mails.tsinghua.edu.cn

Yu Li, Weiwei Zhou, Haoming Dang, Mona Shao, Xiwei Zhao, Zhangang Lin, Jinghe Hu, Jingping Shao
JD.com
liyu1078@jd.com,   zhouweiwei14@jd.com,   danghaoming@jd.com, chewy_lovely@163.com, zhaoxiwei@jd.com, linzhangang@jd.com,   hujinghe@jd.com,   shaojingping@jd.com

CP2

Adaptive Label Smoothing To Regularize Large-Scale Graph Training

Kaixiong Zhou
Rice University
kaixiong.zhou@rice.edu

CP3

Extension of the Dip-Test Repertoire - Efficient and Differentiable P-Value Calculation for Clustering

Over the last decade, the Dip-test of unimodality has gained increasing interest in the data mining community as it is a parameter-free statistical test that reliably rates the modality in one-dimensional samples. It returns a so called Dip-value and a corresponding probability for the sample's unimodality (Dip-p-value). These two values share a sigmoidal relationship. However, the specific transformation is dependent on the sample size. Many Dip-based clustering algorithms use bootstrapped look-up tables translating Dip- to Dip-p-values for a certain limited amount of sample sizes. We propose a specifically designed sigmoid

function as a substitute for these state-of-the-art look-up tables. This accelerates computation and provides an approximation of the Dip- to Dip-p-value transformation for every single sample size. Further, it is differentiable and can therefore easily be integrated in learning schemes using gradient descent. We showcase this by exploiting our function in a novel subspace clustering algorithm called Dip'n'Sub. We highlight in extensive experiments the various benefits of our proposal.

Lena Bauer
University of Vienna
lena.bauer@univie.ac.at

Collin Leiber
LMU Munich
leiber@dbs.ifi.lmu.de

Christian Böhm, Claudia Plant
University of Vienna
christian.boehm@univie.ac.at,   claudia.plant@univie.ac.at

CP3

Beyond The Evidence Lower Bound: Dual Variational Graph Auto-Encoders For Node Clustering

Variational Graph Auto-Encoders have achieved promising performance in several applications. Some recent models incorporate the clustering inductive bias by imposing non-Gaussian prior distributions. However, the regularization term is practically insufficient to learn the clustering structures due to the mismatch between the target and the learned distributions. Thus, we formulate a new variational lower bound that incorporates an explicit clustering objective function. The introduction of a clustering objective leads to two problems. First, the latent information destroyed by the clustering process is critical for generating the between-cluster edges. Second, the noisy and sparse input graph does not benefit from the information learned during the clustering process. To address the first problem, we identify a new term overlooked by existing Evidence Lower BOunds (ELBOs). This term accounts for the difference between the variational posterior used for the clustering task and the variational posterior associated with the generation task. Furthermore, we find that the new term increases resistance to posterior collapse. Theoretically, we demonstrate that our lower bound is a tighter approximation of the log-likelihood function. To address the second problem, we propose a graph update algorithm that reduces the over-segmentation and under-segmentation problems. Our results show that the proposed method improves the clustering quality compared to state-of-the-art VGAE models.

Nairouz Mrabah, Mohamed Bouguessa
University of Quebec at Montreal
mrabah.nairouz@courrier.uqam.ca,
bouguessa.mohamed@uqam.ca

Riadh Ksantini
University of Bahrain
ksantini@uob.edu.bh

CP3

A Lagrangian-Based Approach to Learn Distance Metrics for Clustering with Minimal Data Transformation

Distance metric learning algorithms aim to learn how to measure similarities between data objects in a metric space. In the context of clustering, metric learning typically relies on side-information provided by experts, most commonly expressed in the form of pairwise constraints. In this setting, algorithms for metric learning execute data transformations that bring pairs of data points involved in must-link constraints close together, whereas pair of points involved in cannot-link constraints are moved away from each other. One caveat to such methods is that they can considerably change the original data distribution properties. With that in mind, we propose a Lagrangian-based approach to assist distance metric learning algorithms for clustering. Our method is developed to identify the least impactful transformations to the original data space, while still learning a more suitable metric space for grouping the data using the provided side information. Our results demonstrate that the proposed methodology is able to achieve a competitive clustering performance with respect to truth classification. Furthermore, the method is able to provide more accurate views of the transformed datasets, which can lead to more reliable clustering interpretations.

Rodrigo Randel, Daniel Aloise, Alain Hertz
Polytechnique Montreal
rodrigo.randel@polymtl.ca,   daniel.aloise@polymtl.ca, alain.hertz@polymtl.ca

CP3

Reinforced EM Algorithm for Clustering with Gaussian Mixture Models

Methods that employ the EM algorithm for parameter estimation typically face a notorious yet unsolved problem that the initialization input significantly impacts the algorithm output. We here develop a Reinforced Expectation Maximization (REM) algorithm for cluster analysis using Gaussian mixture models. The competence of REM is achieved by introducing two innovative strategies into the EM framework: (1) a mode-finding strategy for initialization that detects non-trivial modes in the data, and (2) a mode-pruning strategy for detecting true modes/mixture components of the population. The pruning strategy is well-justified in the context of mixture modelling, and we present theoretical guarantees on the quality of the initialization. Extensive experimental studies on both synthetic and real datasets show that our approach achieves better performance compared to state-of-the-art methods.

Joshua Tobin
Trinity College Dublin
tobinjo@tcd.ie

Chin Pang Ho
City University of Hong Kong
clint.ho@cityu.edu.hk

Mimi Zhang
Trinity College Dublin
mimi.zhang@tcd.ie

CP3

Max-Min Diversification with Fairness Constraints: Exact and Approximation Algorithms

Diversity maximization aims to select a diverse and representative subset of items from a large dataset. It is a

fundamental optimization task that finds applications in data summarization, feature selection, web search, recommender systems, and elsewhere. However, in a setting where data items are associated with different groups according to sensitive attributes like sex or race, it is possible that algorithmic solutions for this task, if left unchecked, will under- or over-represent some of the groups. Therefore, we are motivated to address the problem of *max-min diversification with fairness constraints*, aiming to select k items to maximize the minimum distance between any pair of selected items while ensuring that the number of items selected from each group falls within predefined lower and upper bounds. In this work, we propose an exact algorithm based on integer linear programming that is suitable for small datasets as well as a $\frac{1-\varepsilon}{5}$-approximation algorithm for any parameter $\varepsilon \in (0,1)$ that scales to large datasets. Extensive experiments on real-world datasets demonstrate the superior performance of our proposed algorithms over existing ones.

Yanhao Wang
East China Normal University
yhwang@dase.ecnu.edu.cn

Michael Mathioudakis
University of Helsinki
michael.mathioudakis@helsinki.fi

Jia Li
East China Normal University
jiali@stu.ecnu.edu.cn

Francesco Fabbri
Spotify
francescof@spotify.com

CP4

Reliant: Fair Knowledge Distillation for Graph Neural Networks

Graph Neural Networks (GNNs) have shown satisfying performance on various graph learning tasks. To achieve better fitting capability, most GNNs are with a large number of parameters, which makes these GNNs computationally expensive. Therefore, it is difficult to deploy them onto edge devices with scarce computational resources, e.g., mobile phones and wearable smart devices. Knowledge Distillation (KD) is a common solution to compress GNNs, where a light-weighted model (i.e., the student model) is encouraged to mimic the behavior of a computationally expensive GNN (i.e., the teacher GNN model). Nevertheless, most existing GNN-based KD methods lack fairness consideration. As a consequence, the student model usually inherits and even exaggerates the bias from the teacher GNN. To handle such a problem, we take initial steps towards fair knowledge distillation for GNNs. Specifically, we first formulate a novel problem of fair knowledge distillation for GNN-based teacher-student frameworks. Then we propose a principled framework named RELIANT to mitigate the bias exhibited by the student model. Notably, the design of RELIANT is decoupled from any specific teacher and student model structures, and thus can be easily adapted to various GNN-based KD frameworks. We perform extensive experiments on multiple real-world datasets, which corroborates that RELIANT achieves less biased GNN knowledge distillation while maintaining high prediction utility.

Yushun Dong, Binchi Zhang
University of Virginia
yd6eb@virginia.edu, epb6gw@virginia.edu

Yiling Yuan
Beijing University of Posts and Telecommunications
yuanyiling@bupt.edu.cn

Na Zou
Texas A&M University
nzou1@tamu.edu

Qi Wang
Northeastern University
q.wang@northeastern.edu

Jundong Li
University of Virginia
jundong@virginia.edu

CP4

It's PageRank All the Way Down: Simplifying Deep Graph Networks

First developed to rank website relevance, PageRank has become ubiquitous in many areas of graph machine learning including deep learning. We demonstrate that a number of recently published deep graph neural networks are qualitatively equivalent to shallow networks utilizing Personalized PageRank (PPR), and that their performance improvements over existing PPR implementations can be fully explained by hyperparameter choices. We also show that PPR with these hyperparameters outperform more recently published sophisticated variations of PPR-based graph neural networks, and present efficient implementations that reduce training times and memory requirements while improving scalability.

Dominic Jack
University of Melbourne
dom.jack@unimelb.edu.au

Sarah M. Erfani
The University of Melbourne
The University of Melbourne
sarah.erfani@unimelb.edu.au

Jeffrey Chan
RMIT University
jeffrey.chan@rmit.edu.au

Sutharshan Rajasegarar
Deakin University
sutharshan.rajasegarar@deakin.edu.au

Christopher Leckie
The University of Melbourne
caleckie@unimelb.edu.au

CP4

Heterogeneous Graph Contrastive Multi-View Learning

Inspired by the success of Contrastive Learning (CL) in computer vision and natural language processing, Graph Contrastive Learning (GCL) has been developed to learn discriminative node representations on graph datasets. However, the development of GCL on Heterogeneous Information Networks (HINs) is still in the infant stage. For

example, it is unclear how to augment the HINs without substantially altering the underlying semantics, and how to design the contrastive objective to fully capture the rich semantics. Moreover, early investigations demonstrate that CL suffers from sampling bias, whereas conventional debiasing techniques are empirically shown to be inadequate for GCL. How to mitigate the sampling bias for heterogeneous GCL is another important problem. To address the aforementioned challenges, we propose a novel Heterogeneous Graph Contrastive Multi-view Learning (HGCML) model. In particular, we use metapaths as the augmentation to generate multiple subgraphs as multi-views, and propose a contrastive objective to maximize the mutual information between any pairs of metapath-induced views. To alleviate the sampling bias, we further propose a positive sampling strategy to explicitly select positives for each node via jointly considering semantic and structural information preserved on each metapath view. Extensive experiments demonstrate HGCML consistently outperforms state-of-the-art baselines on five real-world benchmark datasets.

Zehong Wang
University of Leeds
mm22zw@leeds.ac.uk

Qi Li, Donghua Yu
Shaoxing University, Shaoxing, China
liqi0713@foxmail.com, donghuayu163@163.com

Xiaolong Han
Nanjing University of Information Science and Technology,
Nanjing, China
202212490770@nuist.edu.cn

Xiao-Zhi Gao
University of Eastern Finland, Kuopio, Finland
xiao-zhi.gao@uef.fi

Shigen Shen
Huzhou University, Huzhou, China
Shaoxing University, Shaoxing, China
shigens@zjhu.edu.cn

CP4
Multimodal Graph Learning for Cross-Modal Retrieval

Cross-modal retrieval has attracted much attention lately for its various applications in Internet data mining. Existing approaches mainly adopt the projection function learning paradigm to construct dual-stream models, which suffer from two limitations: 1) They only utilize the correlations provided by cross-modal data pairs but the multiple correlations among data are unexplored. 2) They typically face the challenge of abstractness of semantics, which means that an instance may have distinct semantic information in different scenarios. In this paper, we propose a novel graph learning based framework termed Multimodal Graph Learning for cross-modal retrieval (MGL), which aims to fully exploit multiple correlations embedded in multimodal data and leverage a graph neural network to capture complementary information to alleviate the information sparsity and abstractness of semantics. First, we propose a graph construction algorithm to explore diverse multimedia information. Second, a modal feature projector is designed to learn modality-shared information, and a co-attention mechanism module is proposed to capture complementary information and perform dynamic feature integration. Third, a fusion and gate module is proposed to fully aggregate captured information and perform denoising. Furthermore, we employ a graph sampling algorithm to make our approach flexible to large-scale scenarios. Experimental results on three benchmark datasets prove the effectiveness of MGL.

Jingyou Xie
Sun Yat-Sen University, Shenzhen, China
xiejy73@mail2.sysu.edu.cn

Zishuo Zhao, Zhenzhou Lin, Ying Shen
Sun Yat-Sen University
zhaozsh@mail2.sysu.edu.cn,   linzhzh6@mail2.sysu.edu.cn,
sheny76@mail.sysu.edu.cn

CP4
Adversarial Hard Negative Generation for Complementary Graph Contrastive Learning

Graph contrastive learning (GCL) has attracted rising research attention recently due to its effectiveness in self-supervised graph learning. A key step of GCL is to conduct data augmentation, based on which self-supervised learning is performed through the contrast between two augmented data views. Existing approaches generally generate the two data views from the original graph, which has been revealed to be less effective due to the lack of data diversity. Meanwhile, although the data augmentation methods and the contrastive modes have been extensively studied, the effect of hard negative samples (i.e.samples that are difficult to distinguish from an anchor node) on GCL is not fully explored. In this paper, we propose a novel complementary graph contrastive learning method boosted by adversarial hard negative sample generation. Specifically, we first construct a kNN graph as the complementary counterpart of the original graph in the semantic space. Then graph augmentation is conducted in both the semantic and topology spaces for the two complementary graphs to obtain two contrastive views with a larger data diversity. To facilitate the contrastive learning, an adversarial network named ADNet is also proposed to generate hard negative samples. The generated samples are more informative and challenging, and thus can further boost the learning performance.

Senzhang Wang, Hao Yan
Central South University
senzhangwang@gmail.com, csuyh1999@csu.edu.cn

Jinlong Du
Nanjing University of Aeronautics and Astronautics
kingloon@.nuaa.edu.cn

Jun Yin
Central South University
8208180806@csu.edu.cn

Junxing Zhu
National University of Defense Technology
zhujunxing@nudt.edu.cn

Chaozhuo Li
Beihang University
lichaozhuo@buaa.edu.cn

Jianxin Wang
Central South University, China

jxwang@csu.edu.cn

CP5

StAGN: Spatial-Temporal Adaptive Graph Network via Contrastive Learning for Sleep Stage Classification

Sleep stage classification is a critical concern in sleep quality assessment and disease diagnosis. Graph network based studies for sleep stages classification have achieved promising performance. However, these studies still ignored the importance of learning morphological feature information with the spatial-temporal relationship among multi-modal physiological signals. To address this issue, we propose a **S**patial-**t**emporal **A**daptive **G**raph **N**etwork named **StAGN** for sleep stage classification. The main advantage of StAGN is to adaptively learn the time-dependent and channel-wise interdependent waveform morphological features in multi-modal physiological signals. Such features will be extracted by a modified 1-dimensional ResNet with a projection shortcut connection and adjusted by a joint spatial-temporal attention, thereby best serving the following brain topological connection graph network for sleep stage classification. Meanwhile, we leverage the contrastive learning scheme with label information to further improve classification accuracy without changing the signal morphology. Experiment results on two publicly available sleep datasets of ISRUC-S1 and ISRUC-S3 show that the proposed StAGN can achieve a competitive performance for sleep stage classification, which is superior to the state-of-the-art counterparts.

Junyang Chen
South China Normal University
School of Computer Science
chenjy@m.scnu.edu.cn

Yidan Dai
School of Computer Science
South China Normal University
daiyidan@m.scnu.edu.cn

Xianhui Chen
Department of Electrical and Computer Engineering
New York University
xhchen@nyu.edu

Yingshan Shen
School of Computer Science
South China Normal University
shenyingshan@scnu.edu.cn

Yan Luximon, Hailiang Wang
School of Design
Hong Kong polytechnic University
yan.luximon@polyu.edu.hk, hailiang.wang@polyu.edu.hk

Yuxin He
College of Big Data and Internet
Shenzhen Technology University
heyuxin@sztu.edu.cn

Wenjun Ma
School of Computer Science
South China Normal University
mawenjun@scnu.edu.cn

Xiaomao Fan

College of Big Data and Internet
Shenzhen Technology University
astrofan2008@gmail.com

CP5

Physics-Guided Meta-Learning Method in Baseflow Prediction over Large Regions

Physics-based groundwater flow equations are powerful tools for water resource assessment under different hydrological and climatic conditions. How these conditions affect the discharge of groundwater (i.e., baseflow) into rivers is one of the most important topics in the hydrology domain. However, due to the different environmental conditions in different basins, it is difficult to use a single physics-based equation to represent the discharge of groundwater in all river basins. Despite the promise of data-driven models in capturing complex relationships, they are also limited in learning heterogeneous baseflow patterns from multiple basins, especially with sparse training data. In this paper, we propose a new data-driven model Physics Guided MeTa Learning (PGMTL), which uses meta-learning to adapt the predictive model to multiple basins and also enhance the meta-learning process with knowledge embodied in different physics-based equations so as to improve the baseflow prediction over a large number of river basins. Experimental results show that our proposed PGMTL has a significant improvement over either physics-based equations or ML models. Moreover, our method has been shown to perform much better with sparse or localized training data. Finally, our method is able to interpret the contribution of each physics-based equation under different scenarios.

Shengyu Chen
University of Pittsburgh, U.S.
shc160@pitt.edu

Yiqun Xie
University of Maryland
xie@umd.edu

Xiang Li
University of Minnesota
lixx5000@umn.edu

Xu Liang
University of Pittsburgh
xuliang@pitt.edu

Xiaowei Jia
U. of Pittsburgh
xiaowei@pitt.edu

CP5

Estimating Latent Population Flows from Aggregated Data via Inversing Multi-Marginal Optimal Transport

We study the problem of estimating latent population flows from aggregated count data. This problem arises when individual trajectories are not available due to privacy issues or measurement fidelity. Instead, the aggregated observations are measured over discrete-time points, for estimating the population flows among states. Most related studies tackle the problems by learning the transition parameters of a time-homogeneous Markov process. Nonetheless, most real-world population flows can be influenced by various uncertainties such as traffic jam and weather conditions.

Thus, in many cases, a time-homogeneous Markov model is a poor approximation of the much more complex population flows. To circumvent this difficulty, we resort to a Multi-marginal Optimal Transport (MOT) formulation that can naturally represent aggregated observations with constrained marginals, and encode time-dependent transition matrices by the cost functions. In particular, we propose to estimate the transition flows from aggregated data by learning the cost functions of the MOT framework, which enables us to capture time-varying dynamic patterns. The experiments demonstrate the improved accuracy of the proposed algorithms than the related methods in estimating several real-world transition flows.

Sikun Yang
Shenzhen Institute of Artificial Intelligence and Robotics for Society; The CUHK, Shenzhen
sikunyang@gmail.com

CP5
STM-GAIL: Spatial-Temporal Meta-GAIL for Learning Diverse Human Driving Strategies

With large amounts of human-generated spatial-temporal urban data (e.g., GPS trajectories of vehicles, passengers trip data on buses and trains, etc.), human urban strategy analysis has become an important problem in many urban scenarios. This problem is hard to solve due to two major challenges: (1) data scarcity and (2) data heterogeneity. Most of the existing works on this problem usually require a large amount of historical observations aiming to correctly infer a human agents urban strategy and thus fail to properly address both challenges at the same time. To solve the human urban strategy analysis problem in case of data scarcity and data heterogeneity, we design a novel learning paradigm Spatial-Temporal Meta-GAIL (STM-GAIL), which can successfully learn diverse human urban strategies from heterogeneous human-generated spatial-temporal urban data. STM-GAIL models the human decision processes as variable length Markov decision processes (VLMDPs) and incorporates the surrounding spatial feature patterns (e.g., traffic volume patterns, etc.) into states to better capture the spatial-temporal dependencies of human decisions. Besides, STM-GAIL learns diverse human urban strategies from the meta-learning perspective, and can be quickly adapted to a new human experts urban strategy with a single trajectory. Extensive experiments on real-world human-generated spatial-temporal dataset are performed.

Yingxue Zhang
Binghamton University
yzhang42@binghamton.edu

Yanhua Li
Worcester Polytechnic Institute
Worcester, MA
yli15@wpi.edu

Xun Zhou
University of Iowa
xun-zhou@uiowa.edu

Ziming Zhang
Worcester Polytechnic Institute
zzhang15@wpi.edu

Jun Luo
Lenovo Group Limited
jluo1@lenovo.com

CP5
Towards Learning in Grey Spatiotemporal Systems: A Prophet to Non-Consecutive Spatiotemporal Dynamics

Spatiotemporal forecasting is an imperative topic in data science due to its critical applications in smart cities. Existing works mostly perform consecutive predictions of following steps with observations continuously obtained, where nearest observations can be exploited as the key knowledge for status estimation. However, the practical issues of early activity planning and sensor failures elicit a new task, non-consecutive forecasting. In this paper, we define spatiotemporal learning systems with missing observations as Grey Spatiotemporal Systems (G2S) and propose a Factor-Decoupled learning framework for G2S to hierarchically decouple multi-level factors, and enable flexible aggregations with uncertainty estimations. We especially select representative sequences to capture periodicity and instantaneous variations, and infer the non-consecutive future statuses under expected exogenous factors, compensating the missing observations. Given the inherent incompleteness and critical applications of G2S, a DisEntangled Uncertainty Quantification is put forward, to identify two types of uncertainty for model interpretations and robustness promotions. Experiments demonstrate that our solution can promote the performance by at least 8.50% on early planning and 2.01%-18.00% on sensor failures.

Zhengyang Zhou, Kuo Yang, Wei Sun, Binwu Wang
University of Science and Technology of China
zzy0929@mail.ustc.edu.cn, yangkuo@mail.ustc.edu.cn,
sunwei3@mail.ustc.edu.cn, wbw1995@mail.ustc.edu.cn

Min Zhou
Huawei Technologies co. ltd
zhoum1900@163.com

Yunan Zong, Yang Wang
University of Science and Technology of China
zyn0728@mail.ustc.edu.cn, angyan@ustc.edu.cn

CP6
Saliency-Augmented Memory Completion for Continual Learning

Continual Learning is considered a key step toward next-generation AI. Among various methods, replay-based approaches that maintain and replay a small episodic memory of previous samples are one of the most successful strategies against catastrophic forgetting. However, since forgetting is inevitable given bounded memory and unbounded tasks, how to forget is a problem continual learning must address. Therefore, beyond simply avoiding (catastrophic) forgetting, an under-explored issue is how to reasonably forget while ensuring the merits of human memory, including 1) storage efficiency, 2) generalizability, and 3) some interpretability. To achieve these simultaneously, our paper proposes a new saliency-augmented memory completion framework for continual learning, inspired by recent discoveries in memory completion/separation in cognitive neuroscience. Specifically, we innovatively propose to store the part of the image most important to the tasks in episodic memory by saliency map extraction and memory encoding. When learning new tasks, previous data from memory are inpainted by an adaptive data generation module, which is inspired by how humans "complete" episodic memory.

The module's parameters are shared across all tasks and it can be jointly trained with a continual learning classifier as bilevel optimization. Extensive experiments on several continual learning and image classification benchmarks demonstrate the proposed method's effectiveness and efficiency.

Guangji Bai, Chen Ling, Yuyang Gao, Liang Zhao
Emory University
guangji.bai@emory.edu, chen.ling@emory.edu, yuyang.gao@emory.edu, liang.zhao@emory.edu

CP6
Ml4c: Seeing Causality Through Latent Vicinity

Haoyue Dai
Carnegie Mellon University
hyda@cmu.edu

CP6
CESED: Exploiting Hyperspherical Predefined Evenly-Distributed Class Centroids for OOD Detection

Out-of-distribution (OOD) detection is critical for ensuring the safe deployment of machine learning models in the open world. Due to the simplicity and intuitiveness of distance-based methods, i.e., samples are detected as OOD if they are relatively far away from the centroids or prototypes of in-distribution (ID) classes, they have attracted widespread attention from researchers in the field of OOD detection. However, prior OOD detection methods directly take off-the-shelf loss functions, like widely used softmax cross-entropy (CE) loss, that suffices for classifying ID samples, but is not optimally designed for OOD detection. In this work, we propose CESED, an improved CE loss applied to the scalable Squared Euclidean Distance vector, which exploits hyperspherical evenly-distributed class centroids for OOD detection. CESED can promote strong ID-OOD separability because it explicitly encourages maximization of inter-class distances and minimization of intra-class distances. Extensive experiments demonstrate that CESED achieves superior detection performance on a comprehensive suite of benchmark datasets. For the more challenging case where CIFAR-100 is used as ID, our method achieves a 31.98% reduction in average FPR95 and 6.20% reduction in ID test error compared to the baseline method using a softmax confidence score. keywords: Image Classification, Out-of-Distribution Detection, Hyperspherical Embedding, Cross-Entropy Loss

Shuai Feng, Wenyu Jiang, Mingcai Chen, Yuntao Du, Hao Cheng, Yuxin Ge, Chongjun Wang
State Key Laboratory for Novel Software Technology
Nanjing University
shuaifeng@smail.nju.edu.cn, lygjwy@smail.nju.edu.cn, chenmc@smail.nju.edu.cn, duyuntao@smail.nju.edu.cn, chengh@smail.nju.edu.cn, yuxinge@smail.nju.edu.cn, chjwang@nju.edu.cn

CP6
An Interpretable Measure of Dataset Complexity for Imbalanced Classification Problems

The class imbalance problem is associated with harmful classification bias and presents itself in a wide variety of important applications of supervised machine learning. Measures have been developed to determine the imbalance complexity of datasets with imbalanced classes. The most common such measure is the Imbalance Ratio (IR). It is, however, widely accepted that the complexity of a classification task is the combined result of class imbalance and other factors, such as class overlap. Thus, in order to accurately assess the complexity of a problem, the data complexity measures ought to account for more than the simple IR. In this paper, we demonstrate that IR has a weak correlation with classifier performance in terms of macro averaged recall, gmean score, and precision. Other more complete measures such as the adapted N1 and N3 measures use neighborhood information to assess overlap. These measures show a strong negative correlation with classifier performance, but their reported values were hard to interpret. This motivates a new measure that estimates overlap complexity and returns a value with a clear interpretation. Here we propose such a measure based on the number of minority instances entangled in a Tomek Link. The proposed measure is evaluated on a large selection of synthetic and real datasets and is found to be as good as or better than the best competitors in terms of its negative correlation with respect to mean classifier performance.

Jonatan Gøttcke
University of Southern Denmark
goettcke@imada.sdu.dk

Colin Bellinger
National Research Council of Canada
colin.bellinger@gmail.com

Paula Branco
University of Ottawa
pbranco@uottawa.ca

Arthur Zimek
University of Southern Denmark,
zimek@imada.sdu.dk

CP6
Uncertainty in Selective Bagging: a Dynamic Bi-Objective Optimization Model

Bagging is a common approach in ensemble learning that generates a group of classifiers through bootstrapping for classification tasks. Despite its wide applications, generating redundant classifiers remains a central challenge in bagging. Many selective bagging models have been presented to deal with this challenge. These models mostly focused on the accuracy of classifiers and their diversity. Despite the importance of uncertainty in the performance of ensemble classifiers, this criterion has been neglected in selective bagging models. In this paper, we propose a two-stage selective bagging model. In the first stage, we formalize the selective bagging problem as a bi-objective optimization model considering both the uncertainty and accuracy of classifiers. We propose an adaptive evolutionary Two-Arch2 algorithm to solve the bi-objective model. The output of this stage is a subset of diverse classifiers which are certain about correct predictions and uncertain about incorrect predictions. While most selective bagging models focus on selecting a fixed subset of classifiers for all test samples, our proposed model has a dynamic approach to the selection process. So, in the second stage of the model, we select only certain classifiers to make an ensemble prediction for each test sample. Experimental results on twenty data sets and comparing with two ensemble models and five state-of-the-art dynamic selective bagging

models show the outperformance of the proposed model.

Mansoureh Maadi, Hadi Akbarzadeh Khorshidi
The University of Melbourne
mmaadi@student.unimelb.edu.au,
hadi.khorshidi@unimelb.edu.au

Uwe Aickelin
The university of Melbourne
uwe.aickelin@unimelb.edu.au

CP7

Heavy Nodes in a Small Neighborhood: Algorithms and Applications

We introduce a weighted and unconstrained variant of the well-known minimum k union problem: Given a bipartite graph $\mathcal{G}(U, V, E)$ with weights for all nodes in V, find a set $S \subseteq V$ such that the ratio between the total weight of the nodes in S and the number of their distinct incident nodes in U is maximized. Our problem, which we term Heavy Nodes in a Small Neighborhood (HNSN), finds applications in marketing, team formation, and money laundering detection. For example, in the latter application, S represents bank account holders who obtain illicit money from some peers of a criminal and route it through their accounts to a target account belonging to the criminal. We prove that HNSN can be solved exactly in polynomial time via linear programming. As the size of $\mathcal{G}$ can be very large in practice, we also develop a near linear-time greedy heuristic. In addition, we formalize a money laundering scenario involving multiple target accounts and show how our algorithms can be extended to deal with it. Our experiments on real and synthetic datasets show that our algorithms find optimal or near-optimal solutions, outperforming a natural baseline, and that they can detect money laundering much more effectively and efficiently than a state-of-the-art method.

Huiping Chen, Grigorios Loukides
King's College London
huiping.chen@kcl.ac.uk, grigorios.loukides@kcl.ac.uk

Robert Gwadera, Solon P. Pissis
CWI Amsterdam
robert.gwadera@epfl.ch, solon.pissis@cwi.nl

CP7

An Index For Temporal Closeness Computation in Evolving Graphs

Temporal closeness is a generalization of the classical closeness centrality measure for analyzing evolving networks. The temporal closeness of a vertex v is defined as the sum of the reciprocals of the temporal distances to the other vertices. Ranking all vertices of a network according to the temporal closeness is computationally expensive as it leads to a single-source-all-destination (SSAD) temporal distance query starting from each vertex of the graph. To reduce the running time of temporal closeness computations, we introduce an index to speed up SSAD temporal distance queries called *Substream* index. We show that deciding if a *Substream* index of a given size exists is NP-complete and provide an efficient greedy approximation. Moreover, we improve the running time of the approximation using min-hashing and parallelization. Our evaluation with real-world temporal networks shows a running time improvement of up to one order of magnitude compared to

the state-of-the-art temporal closeness ranking algorithms.

Lutz Oettershagen, Petra Mutzel
University of Bonn
lutz.oettershagen@cs.uni-bonn.de,   petra.mutzel@cs.uni-bonn.de

CP7

An Efficient Algorithm for Assessing the Number of st-Paths in Large Graphs

Counting the number of subgraphs, or patterns, of a certain kind is at the heart of data mining, and st-paths are one of the most basic graph patterns to express connectivity. The problem of counting the number of st-paths in a graph, both directed and undirected, has been studied since the 70s, and is one of the original #P-complete problems introduced by Valiant [Valiant, SIAM Journal on Computing, 1979]. However, counting can be a heavy task and known algorithms already struggle on graphs with hundreds of nodes. For this reason we propose a novel approach: we assess whether the number of st-paths of an undirected graph is *at least* a given number z. Instead of finding paths one-by-one (i.e., listing), our algorithm is based on decomposing and collapsing computational tasks arranged in a tree-like structure to enhance the effectiveness of each step in growing the number of paths found. Extensive experimental results on real-world datasets show the algorithm scaling to graphs with millions of nodes and edges, with z in the trillions. Its performance is orders of magnitude better than state-of-the-art listing algorithms adapted to this task.

Giulia Punzi, Alessio Conte, Roberto Grossi
Università di Pisa
giuliagpunzi@gmail.com, alessio.conte@unipi.it,
roberto.grossi@unipi.it

Andrea Marino
Universita di Firenze
andrea.marino@unifi.it

CP7

AlignGraph: A Group of Generative Models for Graphs

It is challenging for generative models to learn a distribution over graphs because of the lack of permutation invariance: nodes may be ordered arbitrarily across graphs, and standard graph alignment is combinatorial and notoriously expensive. We propose AlignGraph, a group of generative models that combine fast and efficient graph alignment methods with a family of deep generative models that are invariant to node permutations. Our experiments demonstrate that our framework successfully learns graph distributions, outperforming competitors by $25\% - 560\%$ in relevant performance scores.

Kimia Shayestehfard
Northeastern University
kshayestehfard@ece.neu.edu

Dana H. Brooks
Northeastern University
Dept. of Electrical and Computer Engineering
brooks@ece.neu.edu

Stratis Ioannidis
Northeastern University

ioannidis@ece.neu.edu

CP7
Representation Learning on Dynamic Network of Networks

Network of networks (NoN) where each node of the main network represents a domain-specific network, is a powerful multi-network model to capture the relationships among entities at both coarse and fine granularities. Existing graph convolutional networks (GCN) learn node representations either on a single network or multiple networks while overlooking the relationships among different networks (e.g., main network structure). In addition, many real-world networks often evolve over time, which makes it imperative yet even more challenging to leverage temporal information for node representation learning. In this paper, we study the node representation learning problem on dynamic network of networks. The key idea of designing the static model is the predict-then-propagate strategy such that node representations are obtained by propagating the initial representations of common nodes which are shared across domain-specific networks. To leverage the temporal information underlying dynamic NoN, we extend the static model by a gated recurrent unit (GRU) to capture the dynamics behind cross-network consistency and a self-attention mechanism to learn the dependence of nodes on their historical representations. Armed with these components, we propose an end-to-end model named DRANON to learn node representations on dynamic NoN. We conduct extensive experiments on the dynamic network alignment task, which demonstrate the superior performance of DRANON compared with the state-of-the-arts.

Si Zhang
University of Illinois at Urbana-Champaign
University of Illinois at Urbana-Champaign
sizhang2@illinois.edu

Yinglong Xia, Yan Zhu
Meta
yxia@meta.com, yzhu@meta.com

Hanghang Tong
University of Illinois at Urbana-Champaign
htong@illinois.edu

CP8
A Two-View EEG Representation for Brain Cognition by Composite Temporal-Spatial Contrastive Learning

Electroencephalography (EEG) is a major tool for studying neurophysiological processes. Investigating reliable representations from highly noisy measurements is a pending challenge, however, the medically treasured and insufficient labeled data have driven this process away from a supervised learning manner. Recent works have turned their attention to self-supervised learning (SSL), putting the contrastive strategy on capturing the spatio-temporal characteristics of the neuronal events of interest. We argue that the temporal-spatial view is not the best choice for the SSL contrastive objective because there is a missing piece of the EEG representation that is usually ignored: dynamic fluctuations in brain neurons and the statistical learning of analog/artificial neural networks cannot handle the dynamic characteristics well. This paper proposes a novel two-view contrastive learning framework to refine EEG features from local-global and past-future views. An array of spiking neural networks is embedded to project spatio-temporal features onto the spike sequences to represent the dynamic fluctuation information of EEG. Experimenting with sleep stage classification and prediction of lethal epileptic seizures, we verify the proposal competes favorably against the state-of-the-art methods and offers high-quality features, that is, supervised learning on top of them observes a significant improvement in classification after only one training iteration.

Zheng Chen
Osaka University
chen.zheng.bn1@gmail.com

Lingwei Zhu
University of Alberta
lingweizhu95@gmail.com

Haohui Jia
Nara Institute of Science and Technology
jia.haohui.jc1@is.naist.jp

Takashi Matsubara
Osaka University
matsubara@sys.es.osaka-u.ac.jp

CP8
A Hidden Markov Forest Model for Terrain-Aware Flood Inundation Mapping from Earth Imagery

Flood inundation mapping from Earth imagery plays a vital role in rapid disaster response and national water forecasting. However, the problem is non-trivial due to significant imagery noise and obstacles, complex spatial dependency on 3D terrains, spatial non-stationarity, and high computational cost. Existing machine learning approaches are mostly terrain-unaware and are prone to produce spurious results due to imagery noise and obstacles, requiring significant efforts in post-processing. Recently, several terrain-aware methods were proposed that incorporate complex spatial dependency (e.g., water flow directions on 3D terrains) but they assume that the inferred flood surface level is spatially stationary, making them insufficient for a large heterogeneous geographic area. To address these limitations, this paper proposes a novel spatial learning framework called hidden Markov forest, which decomposes a large heterogeneous area into local stationary zones, represents spatial dependency on 3D terrains via zonal trees (forest), and jointly infers the class map in different zonal trees with spatial regularization. We design efficient inference algorithms based on dynamic programming and multi-resolution filtering. Evaluations on real-world datasets show that our method outperforms baselines and our proposed computational refinement significantly reduces the time cost.

Zhe Jiang, Yupu Zhang
University of Florida
zhe.jiang@ufl.edu, y.zhang1@ufl.edu

Saugat Adhikari
University of Alabama at Birmingham
saugat@uab.edu

Da Yan
The University of Alabama at Birmingham
yanda@uab.edu

Arpan Man Sainju

Middle Tennessee State University
arpan.sainju@mtsu.edu

Xiaowei Jia
U. of Pittsburgh
xiaowei@pitt.edu

Yiqun Xie
University of Maryland
xie@umd.edu

CP8

Time-Delayed Multivariate Time Series Predictions

A major issue with real-time monitoring is to collect complete data. Hardware or software failures, network issues or, more frequently, time delays can disrupt such a collection. This results in having two versions of the same information: one in real-time but with potentially missing data, and the another, albeit complete, is delayed. Many works have studied how to handle missing data for classification and prediction. However, to the best of our knowledge, they do not consider how to leverage the delayed complete data to assist in learning the representation of real-time available data with missing values. This is despite the fact that the delayed complete data contain all the information (e.g., periodicities and trends). In this paper, we propose a framework to enhance the representation learning of the real-time available data by aligning the representation of past real-time but with missing data to that of past delayed but complete data. We test both a distance metric and contrastive learning to achieve this alignment. We implement our framework on a Transformer-based model and experiment it on three datasets. The efficiency of our solution is evaluated against seven baselines and considering four distinct patterns of missing data. Our experiments show that this proposal has a significant improvement in prediction accuracy (5.21% on average) over the baselines.

Hao Niu, Guillaume Habault, Roberto Legaspi
KDDI Research, Inc.
ha-niu@kddi.com, xgu-habault@kddi.com,
ro-legaspi@kddi-research.jp

Chuizheng Meng, Defu Cao
University of Southern California
chuizhem@usc.edu, defucao@usc.edu

Shinya Wada, Chihiro Ono
KDDI Research, Inc.
KDDI Corporation
sh-wada@kddi.com, ci-ono@kddi.com

Yan Liu
University of Southern California
yanliu.cs@usc.edu

CP8

Hierarchical Reinforced Urban Planning: Jointly Steering Region and Block Configurations

With the explosive accumulation of urban geographic, mobile, and IoT service data, AI-assisted automated urban planning, with a goal of configuring land-uses, has become an emerging interdisciplinary topic for smart cities. Existing literature mostly views urban planning as a generative task from the perspective of generating land-use configuration images. We propose a new perspective of formulating urban planning as a hierarchical decision process. That is, given a target region with many geographic blocks, a machine planner firstly selects the optimized urban function portfolios, thereafter, sequentially selects the most appropriate POI for each block based on its urban functions and previously-placed POIs over planning steps. We reformulate this process into a hierarchical reinforcement learning task and develop a novel hierarchical reinforced urban planning framework. This framework includes two components: In region-level configuration, we present an actor-critic based method to overcome the challenge of weak reward feedback in planning the urban functions of regions. In block-level configuration, we propose a single-agent iterative POI allocation strategy to model dependencies between POIs and urban functions, and between current and previous POIs. Finally, we present extensive experimental results on real-world urban data to demonstrate the enhanced performances of the planning as hierarchical decision process perspective and the reinforced planning model.

Pengfei Wang
ICT, Chinese Academy of Science; Alibaba Group;
wpf2106@gmail.com

Daniel Wang
Trinity Preparatory School
wangd23@trinityprep.org

Kunpeng Liu
Portland State University
kunpeng@pdx.edu

Dongjie Wang
University of Central Florida
wangdongjie@knights.ucf.edu

Yuanchun Zhou
Computer Network Information Center,
Chinese Academy of Sciences
zyc@cnic.cn

Leilei Sun
Beihang University
leileisun@buaa.edu.cn

Yanjie Fu
University of Central Florida
yanjie.fu@ucf.edu

CP8

Why Are We Waiting? Discovering Interpretable Models for Predicting Sojourn and Waiting Times

Queueing models explain waiting times, predict sojourn times and help to identify and avoid bottlenecks. Domain experts usually create these models by intensive handcrafting, often resulting in idealized models not fitting the actual process behavior well. Discovering queueing models from data can alleviate this effort, but existing methods do not suffice as they are unable to model complex queueing behaviors. We propose a novel approach to discover queueing models for interpretable waiting time prediction using a rich modeling language to fit complex processes. We formalize the problem in terms of the Minimum Description Length (MDL) principle, by which the best model gives the best lossless compression. The resulting optimization problem is computationally hard, and hence we propose the greedy CueMin algorithm to efficiently find good queueing

models from data. Through an extensive set of experiments including a case study on call center data, we show it discovers inherently interpretable models, which explain and predict behavior of waiting lines better than the state of the art.

Boris Wiegand
SHS - Stahl-Holding-Saar
Saarland University
boris.wiegand@stahl-holding-saar.de

Dietrich Klakow
Saarland University
dietrich.klakow@lsv.uni-saarland.de

Jilles Vreeken
CISPA Helmholtz Center for Information Security
jv@cispa.de

CP9

Physics-Guided Graph Diffusion Network for Combining Heterogeneous Simulated Data: An Application in Predicting Stream Water Temperature

This paper introduces a new method for combining simulated data over different types of nodes in heterogeneous graphs to facilitate predictive learning. Simulation has been widely used in scientific domains to mitigate the need for a large number of observation samples. However, simulated data are often created separately for each type of physical systems while interactions amongst different types of systems remain unexplored. Our method is developed in the context of predicting water temperature in stream networks, which is critical for decision making in water management. In particular, we first develop a graph diffusion network (GDN) to model the interactions amongst stream segments and reservoirs in a heterogeneous graph. We use the GDN model to combine simulated data for both streams and reservoirs in the graph, and use the obtained composite simulations to train the GDN model in a semi-supervised manner. Then the GDN model is further fine-tuned using true observations. Since observation data are often sparse and localized, we further leverage the information from simulations to build a reweighting strategy so as to migitage the discrepancy between training and testing data. Our evaluations in the Delaware River Basin has shown the superiority of the proposed method over multiple baselines using either sparse or localized training data. The proposed GDN model also creates a better composite simulation dataset for heterogeneous graphs.

Shengyu Chen
University of Pittsburgh, U.S.
shc160@pitt.edu

Xiaowei Jia
U. of Pittsburgh
xiaowei@pitt.edu

Can Zheng
University of Pittsburgh
caz51@pitt.edu

Yiqun Xie
University of Maryland
xie@umd.edu

Zhe Jiang
University of Florida
zhe.jiang@ufl.edu

Nasrin Kalanat
University of Pittsburgh
nak168@pitt.edu

CP9

Progress: Prototypical Graph Regression Soft Trees for Molecular Property Prediction

In this work, we propose the novel Prototypical Graph Regression Self-explainable Trees (ProGReST) model, which combines prototype learning, soft decision trees, and Graph Neural Networks. In contrast to other works, our model can be used to address various challenging tasks, including compound property prediction. In ProGReST, the rationale is obtained along with prediction due to the model's built-in interpretability. Additionally, we introduce a new graph prototype projection to accelerate model training. Finally, we evaluate PRoGReST on a wide range of chemical datasets for molecular property prediction and perform in-depth analysis with chemical experts to evaluate obtained interpretations. Our method achieves competitive results against state-of-the-art methods.

Dawid D. Rymarczyk, Daniel Dobrowolski, Tomasz Danel
Jagiellonian University
dawid.rymarczyk@student.uj.edu.pl,
daniel.dobrowolski@student.uj.edu.pl,
tomasz.danel@doctoral.uj.edu.pl

CP9

Anomaly Detection Networks and Fuzzy Control Modules for Energy Grid Management with Q-Learning-Based Decision Making

Renewable energy generation has attracted the interest of researchers, but it is volatile, and management systems are vulnerable to malicious attacks. Therefore, security issues are of paramount importance for energy management systems. In this paper, we propose a secure Q-learning- based energy network management system (SQEMS), which consists of an anomaly detection module, a fuzzy control module to mitigate attacks, and a decision-making module to manage the energy grid. Experimental results show that the proposed anomaly detection module has excellent performance on malicious suppliers attacks (MS), and the fuzzy control module can further mitigate the negative effects of false predictions. The robustness analysis shows the effectiveness, robustness, and transferability in anomaly detection and energy management.

Jia-Hao Syu
National Taiwan University
f08922011@ntu.edu.tw

Jerry Chun-Wei Lin
Western Norway University of Applied Sciences
jerrylin@ieee.org

Philip S. Yu
University of Illinois at Chicago
psyu@cs.uic.edu

CP9

Cisum: Learning Cross-Modality Interaction to Enhance Multimodal Semantic Coverage for Mul-

timodal Summarization

Multimodal summarization (MS) aims to generate a summary from multimodal input. Previous works mainly focus on textual semantic coverage metrics such as ROUGE, which considers the visual content as supplemental data. Therefore, the summary is ineffective to cover the semantics of different modalities. This paper proposes a multitask cross-modality learning framework (CISum) to improve multimodal semantic coverage by learning the cross-modality interaction in the multimodal article. To obtain the visual semantics, we translate images into visual descriptions based on the correlation with text content. Then, the visual description and text content are fused to generate the textual summary to capture the semantics of the multimodal content, and the most relevant image is selected as the visual summary. Furthermore, we design an automatic multimodal semantics coverage metric to evaluate the performance. Experimental results show that CISum outperforms baselines in multimodal semantics coverage metrics while maintaining the excellent performance of ROUGE and BLEU.

Litian Zhang
Beihang University
litianzhang@buaa.edu.cn

Xiaoming Zhang
Beihang Univerisity
yolixs@buaa.edu.cn

Ziming Guo, Zhipeng Liu
Beihang University
guoziming@buaa.edu.cn, lzpeng@buaa.edu.cn

CP9

A Physics-Guided NN-Based Approach for Tropical Cyclone Intensity Estimation

In this paper, a regression neural network pGCNN-TC is designed for estimating tropical cyclone (TC) intensity using satellite images, which is an essential step for TC forecasting. The proposed model employs features derived from convective cores near TC centers, group-equivariant convolution layers that learn rotation equivariant representations from data, and an adjusted MSE loss function to alleviate the underestimation tendency on high intensity TCs. We evaluate our proposed model on a benchmark dataset TCIR, where TCs during 2015-2016 are used as testing cases. Our proposed model outperforms all other start-of-the-art models including NN-based models and objective operational techniques in overall performance. Case studies also suggest that convective core features are especially helpful for high intensity TCs when eye structures are not clear in satellite images.

Ziheng Zhou
Tsinghua University
Zhouzihe18@mails.tsinghua.edu.cn

Ying Zhao
Tsinghua University
CST
yingz@mail.tsinghua.edu.cn

Yiyu Qing
Nanjing University of Information Science and Technology
qingyuyi475@gmail.com

Wenming Jiang, Yihan Wu, Wenguang Chen
Tsinghua University
jwm17@mails.tsinghua.edu.cn,
wuyh22@mails.tsinghua.edu.cn, cwg@tsinghua.edu.cn

CP10

DyFormer: A Scalable Dynamic Graph Transformer with Provable Benefits on Generalization Ability

Transformers have achieved great success in several domains, including Natural Language Processing and Computer Vision. However, its application to real-world graphs is less explored, mainly due to its high computation cost and its poor generalizability caused by the lack of enough training data in the graph domain. To fill in this gap, we propose a scalable Transformer-like dynamic graph learning method named Dynamic Graph Transformer (DyFormer) with spatial-temporal encoding to effectively learn graph topology and capture implicit links. To achieve efficient and scalable training, we propose temporal-union graph structure and its associated subgraph-based node sampling strategy. To improve the generalization ability, we introduce two complementary self-supervised pre-training tasks and show that jointly optimizing the two pre-training tasks results in a smaller Bayesian error rate via an information-theoretic analysis. Extensive experiments on the real-world datasets illustrate that DyFormer achieves a consistent 1%-3% AUC gain (averaged over all time steps) compared with baselines on all benchmarks.

Weilin Cong
Penn State
congweilin95@gmail.com

Yanhong Wu, Yuandong Tian, Mengting Gu, Yinglong Xia, Jason Chen
Meta
yhwu@meta.com, yuandong@meta.com,
gumengting333@gmail.com, yxia@meta.com,
jcchen@fb.com

Mehrdad Mahdavi
Penn State
mzm616@psu.edu

CP10

Optimal Intervention on Weighted Networks via Edge Centrality

Suppose there is a spreading process propagating on a weighted graph. Denote the graph's weight matrix as W. How would we reduce the number of nodes affected during the process? This question appears in recent studies about counterfactual outcomes of implementing edge-weight interventions on mobility networks (Chang et al. (2021)). A practical algorithm to reduce infections is by removing edges with the highest edge centrality, defined as the product of two adjacent nodes' eigenscores (Tong et al. (2012)). In this work, we design edge-weight reduction algorithms on static and time-varying weighted networks with theoretical guarantees. First, we prove that edge centrality equals the gradient of the largest eigenvalue of $WW^\top$ (over W) and generalize the gradient for the largest r eigenvalues of $WW^\top$. Second, we design a Frank-Wolfe algorithm for finding the optimal edge-weight reduction to shrink the largest r eigenvalues of $WW^\top$ under any reduction budget. Third, we extend our algorithm to time-varying networks

with guaranteed optimality. We perform a detailed empirical study to validate our approach. Our algorithm significantly reduces the number of infections compared with existing methods on eleven weighted networks. Further, we illustrate several properties of our algorithm: the benefit of choosing r, fast convergence to the optimum, and a linear-scale runtime per iteration.

Dongyue Li, Tina Eliassi-Rad, Hongyang Zhang
Northeastern University
li.dongyu@northeastern.edu,
t.eliassirad@northeastern.edu,
ho.zhang@northeastern.edu

CP10
STABLE: Identifying and Mitigating Instability in Embeddings of the Degenerate Core

Are the embeddings of a graph's degenerate core stable? What happens to the embeddings of nodes in the degenerate core as we systematically remove periphery nodes (by repeatedly peeling off k-cores)? We discover three patterns w.r.t. instability in degenerate-core embeddings across a variety of popular graph embedding algorithms and datasets. We correlate instability with an increase in edge density, and then theoretically show that in the case of ErdosRnyi graphs embedded with Laplacian Eigenmaps, the best and worst possible embeddings become less distinguishable as density increases. Furthermore, we present the STABLE algorithm, which takes an existing graph embedding algorithm and makes it stable. We show the effectiveness of STABLE in terms of making the degenerate-core embedding stable and still producing state-of-the-art link prediction performance.

David M. Liu, Tina Eliassi-Rad
Northeastern University
liu.davi@northeastern.edu, tina@eliassi.org

CP10
Gist: Graph Inference for Structured Time Series

Machine learning and data analytics tasks on graphs enjoy a lot of attention from both researchers and practitioners due to the utility that a graph structure among data entities adds for downstream tasks. In many cases a graph structure is not known a priori, and has to be inferred from data. Specifically, learning a graph associating time series may elucidate hidden dependencies and also enable improved performance in tasks like classification, forecasting and clustering. Approaches based on correlation and precision matrix have been employed widely, recent approaches that model observations as signals on graphs have been shown to be more advantageous. We propose to learn a graph among time series based on similarity of encoding via temporal dictionaries. The key premise is that observed time series have an inherent underlying structure such as periodicity and/or trends and can be succinctly encoded via an appropriate dictionary. Time series with similar encodings are associated via edges in the inferred graph. We formulate the problem as a joint graph Laplacian learning and sparse dictionary-based coding. We consider two alternative solutions for different problem settings: one that associates time series that behave similarly and one that associates them based on shared periodicity. We demonstrate that our solutions enable improved performance over baselines in identifying ground truth edges and groupings of the time series in 8 real-world datasets from diverse do-

mains.

Boya Ma
University at Albany, SUNY
bma@albany.edu

Maxwell J. McNeil
State University of New York, Albany, U.S.
mmcneil2@albany.edu

Petko Bogdanov
University at Albany
pbogdanov@albany.edu

CP10
Node Ranking in Labeled Networks

The entities in directed networks arising from real-world interactions are often naturally organized under some hierarchical structure. Given a directed, weighted, graph with edges and node labels, we introduce ranking problem where the obtained hierarchy should be described using node labels. Such method has the advantage to not only rank the nodes but also provide an explanation for such ranking. To this end, we define a binary tree called label tree, where each leaf represents a rank and each non-leaf contains a single label, which is then used to partition, and consequently, rank the nodes in the input graph. We measure the quality of trees using agony score, a penalty score that penalizes the edges from higher ranks to lower ranks based on the severity of the violation. We show that the problem is NP-hard, and even inapproximable if we limit the size of the label tree. Therefore, we resort to heuristics, and design a divide-and-conquer algorithm which runs in $O((n + m) \log n + \ell R)$, where R is the number of node-label pairs in the given graph, ℓ is the number of nodes in the resulting label tree, and n and m denote the number of nodes and edges respectively. We also report an experimental study that shows that our algorithm can be applied to large networks, that it can find ground truth in synthetic datasets, and can produce explainable hierarchies in real-world datasets.

Chamalee N. Wickrama Arachchi
University of Helsinki
Finland
chamalee.wickramaarachch@helsinki.fi

Nikolaj Tatti
University of Helsinki
nikolaj.tatti@helsinki.fi

CP11
Attention-Based Multi-Modal Missing Value Imputation for Time Series Data with High Missing Rate

Multivariate time series data is prone to a high missing rate which presents an obstacle to statistical analysis of the data. Imputation has become the standard measure to handle this challenge. However, existing time series missing value imputation methods are mostly uni-modal that relies on self-imputation. With an unprecedented rate of data collection, the availability of multi-modal data is increasing, allowing us the opportunity to impute the time series missing values using other datasets generated from the same cohort. In this paper, we propose a multi-modal time series missing value imputation framework, TSEst, that can utilize multiple data modalities to overcome the

limitations of self-imputation. The framework uses additional cross-sectional or time series data for the imputation and therefore, is less affected by a high missing rate in the time series data. A comprehensive set of experiments on two datasets shows an improvement in imputation accuracy over the baselines. Experimental results also demonstrate that the improvement is caused by the effective integration of the additional data modality. The proposed framework can impute missing values in the samples with no time series data available, reducing the reliance on long-term data collection.

Khandakar Tanvir Ahmed, Sudipto Baul, Yanjie Fu, Wei Zhang
University of Central Florida
t.ahmed@knights.ucf.edu, sudipto@knights.ucf.edu, yanjie.fu@ucf.edu, wzhang.cs@ucf.edu

CP11
Exact and Heuristic Approaches to Speeding Up the MSM Time Series Distance Computation

The computation of the distance of two time series is time-consuming for any elastic distance function that accounts for misalignments. Among those functions, DTW is the most prominent. However, a recent extensive evaluation has shown that the move-split merge (MSM) metric is superior to DTW regarding the analytical accuracy of the 1-NN classifier. Unfortunately, the running time of the standard dynamic programming algorithm for MSM distance computation is $\Omega(n^2)$, where n is the length of the longest time series. In this paper, we provide approaches to reducing the cost of MSM distance computations by using lower and upper bounds for early pruning paths in the underlying dynamic programming table. For the case of one time series being a constant, we present a linear-time algorithm. In addition, we propose new linear-time heuristics and adapt heuristics known from DTW to computing the MSM distance. One heuristic employs the metric property of MSM and the previously introduced linear-time algorithm. Our experimental studies demonstrate substantial speed-ups in our approaches compared to previous MSM algorithms. In particular, the running time for MSM is faster than a state-of-the-art DTW distance computation for a majority of the popular UCR data sets.

Jana Holznigenkemper
University of Marburg
holznigenkemper@mathematik.uni-marburg.de

Christian Komusiewicz
Philipps-Universität Marburg
komusiewicz@mathematik.uni-marburg.de

Bernhard Seeger
University of Marburg
seeger@mathematik.uni-marburg.de

CP11
Debiased Imitation Learning for Modulated Temporal Point Processes

Temporal event sequences associated with different event types (e.g., location indices, disease types) are observed in various applications such as disaster resilience, criminology, and healthcare. Temporal point processes (TPPs) have been developed to capture the exciting patterns between events and forecast future events quantitatively. Unfortunately, the events with different types often suffer from unknown biased observations in real-world scenarios due to external interference. Accordingly, the temporal point processes learned by conventional maximum likelihood estimation (MLE) from such biased data may be misspecified and may lead to inaccurate predictions. To overcome this issue, we model biased event sequences as modulating TPPs with additional unknown thinning processes. Furthermore, we develop a novel debiased imitation learning framework to learn the modulated TPPs and suppress the negative influences of biased data, which is more robust than conventional MLE. When applying the debiased imitation learning framework, we design a simple but effective reward function based on the historical embedding obtained by the TPP model. Experiments on three real-world datasets demonstrate that our proposed method significantly outperforms existing methods.

Zhuoqun Li, Zihan Zhou, Mingxuan Sun
Louisiana State University
zli82@lsu.edu, clairechou1112@gmail.com,
msun11@lsu.edu

Hongteng Xu
Renmin University Of China
hongtengxu313@gmail.com

CP11
Spatiotemporal Classification with Limited Labels Using Constrained Clustering for Large Datasets

Creating separable representations via representation learning and clustering is critical in analyzing large unstructured datasets with only a few labels. Separable representations can lead to supervised models with better classification capabilities and additionally aid in generating new labeled samples. Most unsupervised and semisupervised methods to analyze large datasets do not leverage the existing small amounts of labels to get better representations. In this paper, we propose a spatiotemporal clustering paradigm that uses spatial and temporal features combined with a constrained loss to produce separable representations. We show the working of this method on the newly published dataset ReaLSAT, a dataset of surface water dynamics for over 680,000 lakes across the world, making it an essential dataset in terms of ecology and sustainability. Using this large unlabelled dataset, we first show how a spatiotemporal representation is better compared to just spatial or temporal representation. We then show how we can learn even better representations using a constrained loss with few labels. We conclude by showing how our method, using few labels, can pick out new labeled samples from the unlabeled data, which can be used to augment supervised methods leading to better classification.

Praveen Ravirathinam
University of Minnesota, Twin Cities
pravirat@umn.edu

Rahul Ghosh
University of Minnesota
ghosh128@umn.edu

Ke Wang, Keyang Xuan
University of Minnesota, Twin Cities
wan00802@umn.edu, xuan0008@umn.edu

Ankush Khandelwal
University of Minnesota

khand035@umn.edu

Hilary Dugan, Paul Hanson
University of Wisconsin-Madison
hdugan@wisc.edu, pchanson@wisc.edu

Vipin Kumar
University of Minnesota
kumar001@umn.edu

CP11

Probabilistic Decomposition Transformer for Time Series Forecasting

Time series forecasting is crucial for many fields, such as disaster warning, weather prediction, and energy consumption. The Transformer-based models are considered to have revolutionized the field of time series. However, the autoregressive form of the Transformer introduces cumulative errors in the inference stage. Furthermore, the complex temporal pattern of the time series leads to an increased difficulty for the models in mining reliable temporal dependencies. In this paper, we propose the Probabilistic Decomposition Transformer model, which provides a flexible framework for hierarchical and decomposable forecasts. The hierarchical mechanism utilizes the forecasting results of Transformer as conditional information for the generative model, performing sequence-level forecasts to approximate the ground truth, which can mitigate the cumulative error of the autoregressive Transformer. In addition, the conditional generative model encodes historical and predictive information into the latent space and reconstructs typical patterns from the latent space, such as seasonality and trend terms. The process provides a flexible framework for the separation of complex patterns through the interaction of information in the latent space. Extensive experiments on several datasets demonstrate the effectiveness and robustness of the model, indicating that it compares favorably with the state-of-the-art.

Junlong Tong, Liping Xie, Kanjian Zhang
Southeast University
jltong1009@gmail.com, lpxie@seu.edu.cn, kjzhang@seu.edu.cn

CP12

Penalized Non-Linear Canonical Correlation Analysis for Ordinal Data with Application to the International Classification of Functioning, Disability and Health

A non-linear version of canonical correlation analysis (CCA) as an exploratory, data mining technique is proposed that is particularly suited for discrete and ordinal data such as rating scales. The method can be seen as a modification of optimal scaling, while on the one hand quantifications are allowed to be non-monotone and vary across components, and on the other hand the estimated quantifications are penalized such that in the extreme case of maximum penalty the method is equivalent to classical, linear CCA using the original scale. In contrast to kernel and/or deep CCA, the non-linear and potentially non-monotone transformations of the original data are estimated in an explicit yet feature-wise and smoothed fashion, which (a) facilitates interpretation and (b) makes it promising for moderately sized data sets, too. Specifically, we employ an iterated, penalized least-squares algorithm with internal quadratic programming. The method is illustrated and evaluated on synthetic, and more importantly, real

world data coming from an application of the 'International Classification of Functioning, Disability and Health' (ICF). Particularly, it is observed that non-monotone transformations of some of the so-called 'environmental factors' indeed result in substantially higher correlations with other ICF categories than standard, linear CCA.

Jan Gertheiss
Helmut Schmidt University
jan.gertheiss@hsu-hh.de

Russell Shinohara
Department of Biostatistics and Epidemiology
University of Pennsylvania
rshi@mail.med.upenn.edu

CP12

Movae: A Variational AutoEncoder for Molecular Graph Generation

Zerun Lin
Shenzhen University
2060432004@email.szu.edu.cn

CP12

Knowledge-Enhanced Semi-Supervised Federated Learning for Aggregating Heterogeneous Lightweight Clients in IoT

Federated learning (FL) enables multiple clients to train models collaboratively without sharing local data, which has achieved promising results in different areas, including the Internet of Things (IoT). However, end IoT devices do not have abilities to automatically annotate their collected data, which leads to the label shortage issue at the client side. To collaboratively train an FL model, we can only use a small number of labeled data stored on the server. This is a new yet practical scenario in federated learning, i.e., labels-at-server semi-supervised federated learning (SemiFL). Although several SemiFL approaches have been proposed recently, none of them can focus on the personalization issue in their model design. IoT environments make SemiFL more challenging, as we need to take device computational constraints and communication cost into consideration simultaneously. To tackle these new challenges together, we propose a novel SemiFL framework named pFedKnow. pFedKnow generates lightweight personalized client models via neural network pruning techniques to reduce communication cost. Moreover, it incorporates pre-trained large models as prior knowledge to guide the aggregation of personalized client models and further enhance the framework performance. Experiment results on both image and text datasets show that the proposed pFedKnow outperforms state-of-the-art baselines as well as reducing considerable communication cost.

Fenglong Ma, Jiaqi Wang
Pennsylvania State University
fenglong@psu.edu, jqwang@psu.edu

Shenglai Zeng
University of Electronic Science and Technology of China
shenglaizeng@gmail.com

Zewei Long
University of Illinois Urbana-Champaign
zeweil2@illinois.edu

Yaqing Wang
Purdue University
wang5075@purdue.edu

Houping Xiao
Georgia State University
hxiao@gsu.edu

CP12

Extrinsic-Intrinsic Representation Learning Framework for Drug Discovery

Exploring drug-target interaction remains one of the essential tasks in drug discovery, and it is critical to gain a thorough understanding of the biological process and disease mechanisms. Despite recent successes in the application of machine learning approaches, drug-target interaction studies are still largely under-explored due to significant challenges in modeling different types of representations and capturing the inherent correlation between targets and drugs from low-level representations. What is more, the length of the target protein sequences and the complexity of the drug-target binding complex make the problem hard to handle. In this work, we focus on increasing the generalizability and interpretability of the drug-target prediction models and propose an Extrinsic-Intrinsic Representation learning model (EIR) intended to discover the inner correlation between target proteins and drugs on both the extrinsic and intrinsic levels. Our experimental results show that EIR makes more accurate predictions than the state-of-the-art method in both drug-target affinity prediction and drug-target interface prediction tasks and demonstrate the potential of the structural-free method for drug discovery.

Tian Xia, Sarp Aykent, Wei-Shinn Ku
Auburn University
tzx0015@auburn.edu,   sarp@auburn.edu,
weishinn@auburn.edu

CP12

A Multi-Scale Interaction Motion Network for Action Recognition Based on Capsule Network

Recently, action recognition has achieved impressive performance, mainly due to the aid of deep convolutional neural networks and large datasets. Traditionally, most efforts in action recognition have focused on capturing motion information by dense optical flow, but optical flow extraction is very time-consuming. Moreover, prior arts seek to improve accuracy but neglect the part-whole relationship between objects in videos. To this end, we present a novel collaborative multipath capsule network (CMCN) for action recognition. In particular, we propose a plug-and-play collaborative multipath block containing spatiotemporal, channel, and motion units, which are complementary and crucial information for action recognition. We exploit the interaction of these three units and selectively emphasize informative spatial-temporal motion to reduce the expensive computational costs. Subsequently, we explore a new capsule voting procedure to reduce the computation used in the capsule dynamic routing mechanism. The critical insight is that the same type of capsules simulates the same entity in different positions, and their voting results should be consistent. This strategy lessens the number of learning parameters that backward pass in the training process, and thus strengthens part-whole relationships in a video. Extensive experiments on multiple real-world datasets for action recognition demonstrate that our model significantly

outperforms state-of-the-art models.

Xiangping Zheng, Xun Liang, Bo Wu
Renmin University of China
xpzheng@ruc.edu.cn, xliang@ruc.edu.cn,
wubochn@ruc.edu.cn

Jun Wang
Swinburne University of Technology
junwang@swin.edu.au

Yuhui Guo, Xuan Zhang
Renmin University of China
yhguo@ruc.edu.cn, zhangxuanalex@ruc.edu.cn

Yuefeng Ma
Qufu Normal University
yfmqnu@163.com

CP13

Concept Discovery for Fast Adaptation

The advances in deep learning have enabled machine learning methods to outperform human beings in various areas, but it remains a great challenge for a well-trained model to quickly adapt to a new task. One promising solution to realize this goal is through meta-learning, also known as learning to learn, which has achieved promising results in few-shot learning. However, current approaches are still enormously different from human beings' learning process, especially in the ability to extract structural and transferable knowledge. This drawback makes current meta-learning frameworks non-interpretable and hard to extend to more complex tasks. We tackle this problem by introducing concept discovery to the few-shot learning problem, where we achieve more effective adaptation by meta-learning the structure among the data features, leading to a composite representation of the data. Our proposed method Concept-Based Model-Agnostic Meta-Learning (COMAML) has been shown to achieve consistent improvements in the structured data for both synthesized datasets and real-world datasets.

Shengyu Feng
Carnegie Mellon University
Language Technology Institute
shengyuf@andrew.cmu.edu

Hanghang Tong
University of Illinois at Urbana-Champaign
htong@illinois.edu

CP13

Robust Learning via Golden Symmetric Loss of (un)Trusted Labels

Learning robust deep models against noisy labels becomes ever critical when today's data is commonly collected from open platforms and subject to adversarial corruption. The information on the label corruption process, i.e., corruption matrix, can greatly enhance the robustness of deep models but still fall behind in combating hard classes. In this paper, we propose to construct a golden symmetric loss (GSL) based on the estimated corruption matrix as to avoid overfitting to noisy labels and learn effectively from hard classes. GSL is the weighted sum of the corrected regular cross entropy and reverse cross entropy. By leveraging a small fraction of trusted clean data, we estimate the cor-

ruption matrix and use it to correct the loss as well as to determine the weights of GSL. We theoretically prove the robustness of the proposed loss function in the presence of dirty labels. We provide a heuristics to adaptively tune the loss weights of GSL according to the noise rate and diversity measured from the dataset. We evaluate our proposed golden symmetric loss on both vision and natural language deep models subject to different types of label noise patterns. Empirical results show that GSL can significantly outperform the existing robust training methods on different noise patterns, showing accuracy improvement up to 18% on CIFAR-100 and 1% on real world noisy dataset of Clothing1M.

Amirmasoud Ghiassi
Delft University of Technology
s.ghiassi@tudelft.nl

Robert Birke
University of Torino
robert.birke@unito.it

Lydia Chen
Delft University of Technology
lydiaychen@ieee.com

CP13
Adaptive Precision Training (AdaPT): A Dynamic Quantized Training Approach for DNNs

Quantizing deep neural networks (DNNs) is an important strategy for training or inference in time-critical applications. State-of-the-art quantization approaches focus on post-training quantization. While work on quantization during training exists, most approaches require refinement in full precision in the final training phase, use a coarse quantization leading to accuracy loss, or enforce a global bit-width across the entire DNN. This leads to suboptimal assignments of bit-widths to layers and, consequently, suboptimal resource usage. To overcome such limitations, we introduce AdaPT, a new fixed-point quantized training strategy for DNNs. AdaPT decides about precision switches between epochs based on an information theory motivated heuristic. On a per-layer basis, AdaPT chooses the lowest precision that causes no quantization-induced information loss, while keeping the precision high enough such that future learning steps do not suffer from vanishing gradients. The benefits of this quantization are evaluated based on an analytical performance model. We illustrate an average $1.31\times$ (or $4.76\times$ adjusted for iso-accuracy) speedup compared to training in float32 at iso-accuracy, even achieving an average accuracy increase of 0.74 percentage points for AlexNet/ResNet-20 on CIFAR10/CIFAR100/EMNIST and LeNet-5/MNIST. AdaPT trained models reach an average inference $2.28\times$ speedup and a size reduction up to 51% of the corresponding unquantized model.

Lorenz Kummer
University of Vienna
Faculty of Computer Science
lorenz.kummer@univie.ac.at

Kevin Sidak
University of Vienna
kevin.sidak@univie.ac.at

Tabea Reichmann
University of Vienna Faculty of Computer Science
tabea.reichmann@univie.ac.at

Wilfried N. Gansterer
Department of Computer Science
University of Vienna
wilfried.gansterer@univie.ac.at

CP13
BDA: Bandit-Based Transferable AutoAugment

AutoAugment is an automatic method to design data augmentation policies for deep learning, and has achieved significant improvements on computer vision tasks. However, since early AutoAugment approaches cost thousands of GPU hours, there is a recent demand to investigate low-cost search methods that can still find effective augmentation policies. In this paper, we propose a multi-armed bandit algorithm, named Bandit Data Augment (BDA), to efficiently search for optimal and transferable data augmentation policies. We leverage Successive Halving to make the bandit model progressively focus on more promising augmentation operations during the search, leading to sparse selection of operations and more generalizable augmentation policies. We also propose a computationally efficient rewarding scheme to reduce the evaluation cost of augmentation policies. Extensive experiments demonstrate that BDA can achieve comparable or better performance than prior AutoAugment methods on a wide range of models on CIFAR-10/100 and ImageNet benchmarks. Besides, BDA is 555 times and 536 times faster than AutoAugment on CIFAR-10 and ImageNet, respectively. In addition, BDA is 16 times faster than Fast AutoAugment on ImageNet. More importantly, BDA can discover policies that are transferable across datasets and models, and achieve similar performance to policies found directly on the target dataset.

Shan Lu, Mingjun Zhao
University of Alberta
slu1@ualberta.ca, zhao2@ualberta.ca

Songling Yuan, Xiaoli Wang, Lei Yang
Tencent
laneyuan@tencent.com, evexlwang@tencent.com,
yangalalei@gmail.com

Di Niu
University of Alberta
dniu@ualberta.ca

CP13
Domain Disentangled Meta-Learning

A key challenge with supervised learning (e.g., image classification) is the shift of data distribution and domain from training to testing datasets, so-called domain shift (or distribution shift), which usually leads to a reduction of model accuracy. Various meta-learning approaches have been proposed to prevent the accuracy loss by learning an adaptable model with training data, and adapting it to test time data from a new data domain. However, when the domain shift occurs in multiple domain dimensions (e.g., images may be transformed by rotations, transitions, and expansions), the average predictive power of the adapted model will deteriorate. To tackle this problem, we propose a domain disentangled meta-learning (DDML) framework. DDML disentangles the data domain by dimensions, learns the representations of domain dimensions independently, and adapts

to the domain of test time data. We evaluate our DDML on image classification problems using three datasets with distribution shifts over multiple domain dimensions. Comparing to various baselines in meta-learning and empirical risk minimization, our DDML approach achieves consistently higher classification accuracy with the test time data. These results demonstrate that domain disentanglement reduces the complexity of the model adaptation, thus increases the model generalizability, and prevents it from overfitting.

Xin Zhang
Worcester Polytechnic Institute
xzhang17@wpi.edu

Yanhua Li
Worcester Polytechnic Institute
Worcester, MA
yli15@wpi.edu

Ziming Zhang
Worcester Polytechnic Institute
zzhang15@wpi.edu

Zhi-Li Zhang
University of Minnesota
Computer Science
zhzhang@cs.umn.edu

CP14
Mc-Sq: A Highly Accurate Ensemble for Multi-Class Quantification

Quantification research proposes methods to estimate the class distribution in an independent sample. Many areas, such as epidemiology, sentiment analysis, political research and ecological surveillance, rely on quantification methods to estimate aggregated quantities. For instance, epidemiologists are often concerned with the dynamics of the number of disease cases across space and time. Thus, while classification predicts individual subjects, quantification is the class of methods that directly estimate the number of cases. Quantification is a thriving research area, and the community has proposed several approaches in the last decade. Nevertheless, most quantification research has focused on binary-class quantifiers, expecting these approaches to extend to multi-class using the one-versus-all (OVA) approach. However, there is enough empirical evidence indicating the performance of OVA multi-class quantifiers is subpar. This paper has two main contributions. First, we demonstrate why OVA quantifiers are doomed to underperform in multi-class settings due to a distribution shift they cannot handle. Second, we propose a new class of quantifiers based on ensemble learning that boosts the performance of the base quantifiers in the binary and, more importantly, multi-class settings. We show that our ensembles are the best-performing quantifiers compared with 33 state-of-the-art (single and ensemble) quantifiers and rank first in a recent quantification competition.

Gustavo Batista
University of Sao Paulo
g.batista@unsw.edu.au

Zahra Donyavi
University of New South Wales
z.donyavi@unsw.edu.au

Adriane Serapião

São Paulo State University
University of New South Wales
adriane.serapiao@unesp.br

CP14
Optimal Smooth Approximation for Quantile Matrix Factorization

Matrix Factorization (MF) is a critical technique in many applications. Most existing MF methods minimize the L_2 loss between observations and their dependent matrix measurement variables. Under certain conditions, linear convergence to global optimality can be established for L_2 loss, while L_1 loss and check loss are widely used to deal with data that are contaminated with outliers, there lacks of efficient and theoretically proven algorithms specifically designed for MF with non-smooth losses. In this paper, we study Quantile Matrix Factorization (QMF) that adopts a tunable check loss and can introduce robustness to matrix estimation under possibly skewed and heavy-tailed observations, which are prevalent in reality. To deal with the non-smooth loss, we propose Nesterov-smoothed QMF (NsQMF), extending Nesterov's optimal smooth approximation technique to MF. We then present an alternating minimization algorithm to solve NsQMF efficiently while handling non-convexity. We prove that solving the smoothed NsQMF is equivalent to solving the original non-smooth QMF problem and that our algorithm achieves linear convergence to the global optimality of QMF. Extensive evaluations verify our theoretical findings and demonstrate that NsQMF significantly outperforms commonly used Least Squares Matrix Factorization (LSMF) and prior rough smoothing techniques for QMF under various noise distributions.

Peng Liu
University of Kent
pl325@kent.ac.uk

Yi Liu
University of Alberta
yliu16@ualberta.ca

Rui Zhu
Deepmind
rzhu3@ualberta.ca

Linglong Kong, Bei Jiang, Di Niu
University of Alberta
lkong@ualberta.ca, bei1@ualberta.ca, dniu@ualberta.ca

CP14
Multi-Task Learning with Prior Information

Multi-task learning aims to boost the generalization performance of multiple related tasks simultaneously by leveraging information contained in those tasks. In this paper, we propose a multi-task learning framework, where we utilize prior knowledge in the relations between features. We also impose a penalty on the coefficients changing for each specific feature to ensure related tasks have similar coefficients on common features shared among them. In addition, we capture a common set of features via group sparsity. The objective is formulated as a non-smooth convex optimization problem, which can be solved with various methods, including gradient descent method with fixed stepsize, iterative shrinkage-thresholding algorithm (ISTA) with back-tracking, and its momentum variation – fast iter-

ative shrinkage-thresholding algorithm (FISTA). In light of the sub-linear convergence rate of the methods aforementioned, we propose an asymptotically linear convergent algorithm with theoretical guarantee. Empirical experiments on both regression and classification tasks with real-world datasets demonstrate that our proposed algorithms are capable of improving the generalization performance of multiple related tasks.

Mengyuan Zhang, <u>Kai Liu</u>
Computer Science Division
Clemson University
mengyuz@clemson.edu, liukaizhijia@gmail.com

CP14

Learning Causal Structure on Mixed Data with Tree-Structured Functional Models

Discovering causal relations from observational data is at the heart of scientific research. Most causal discovery methods assume that the data have only one variable type. In real-world problems, however, data can consist of a mixture of continuous, discrete, and categorical variables. In this paper, we examine the causal discovery problem on mixed data. We introduce a general tree-structured functional causal model, which is well suited for characterizing the generating mechanisms of mixed data by allowing non-differentiability and nonlinearity. We present corresponding identifiability results, showing that under mild conditions, the causal directions can be uniquely determined from observational distributions. Further, we prove that the causal direction between continuous and discrete variables is generally identifiable under a much larger function class. Based on the theoretical findings, we propose an effective causal discovery method leveraging a consistent score function and powerful tree-learning techniques. Experiments on both synthetic and real data verify the effectiveness of our approach.

Tian Qin, Tian-Zuo Wang, Zhi-Hua Zhou
Nanjing University
qint@lamda.nju.edu.cn,   wangtz@lamda.nju.edu.cn,
zhouzh@nju.edu.cn

CP14

Ranking with Submodular Functions on the Fly

Maximizing submodular functions have been studied extensively for a wide range of subset-selection problems. However, much less attention has been given to the role of submodularity in sequence-selection and ranking problems. A recently-introduced framework, named maximum submodular ranking (MSR), tackles a family of ranking problems that arise naturally when resources are shared among multiple demands with different budgets. For example, the MSR framework can be used to rank web pages for multiple user intents. In this paper, we extend the MSR framework in the streaming setting. In particular, we consider two different streaming models and we propose practical approximation algorithms. In the first streaming model, called function arriving, we assume that submodular functions (demands) arrive continuously in a stream, while in the second model, called item arriving, we assume that items (resources) arrive continuously. Furthermore, we study the MSR problem with additional constraints on the output sequence, such as a matroid constraint that can ensure fair exposure among items from different groups. These extensions significantly broaden the range of problems that can be captured by the MSR framework. On the

practical side, we develop several novel applications based on the MSR formulation, and empirically evaluate the performance of the proposed methods.

Guangyi Zhang
KTH Royal Institute of Technology
guaz@kth.se

Nikolaj Tatti
University of Helsinki
nikolaj.tatti@helsinki.fi

Aristides Gionis
KTH Royal Institute of Technology
argioni@kth.se

CP15

Balancing Task Coverage and Expert Workload in Team Formation

In the classical team-formation problem the goal is to identify a team of experts such that the skills of these experts cover all the skills required by a given task. In this paper, we deviate from this setting and propose a variant of the classical problem in which we aim to cover the skills of every task as well as possible, while also trying to minimize the maximum workload among the experts. Instead of setting the coverage constraint and minimizing the maximum load, we combine these two objectives into one. We call the corresponding assignment problem the Balanced Coverage problem, and show that it is NP-hard. We note that the objective function, which may also take negative values, does not allow us to design approximation algorithms with multiplicative guarantees. Consequently, we adopt a weaker notion of approximation and we show that under this notion we can design a polynomial-time approximation algorithm with provable guarantees. We also describe a set of computational speedups that we can apply to the algorithm to make it scale for reasonably large datasets. From the practical point of view, we demonstrate how the nature of the objective function allows us to efficiently tune the two parts of the objective and tailor their importance to a particular application. Our experiments with a variety of real datasets demonstrate the utility of our problem formulation as well as the efficacy and efficiency of our algorithm in practice.

<u>Karan Vombatkere</u>, Evimaria Terzi
Boston University
kvombat@bu.edu, evimaria@bu.edu

CP15

Taxonomy-Guided Fine-Grained Entity Set Expansion

Entity set expansion, the task of expanding a small set of similar entities into a much larger set, is a vital step for downstream tasks such as named entity recognition, knowledge base construction and information retrieval. Existing entity set expansion methods were developed by mainly considering entities at coarse-grained levels, which encounter difficulties for entity set expansion at fine-grained levels, due to the subtlety on fine-grained type inference and semantic drifting. In this study, we propose an automated (i.e. without human annotation), fine-grained set expansion framework, FGExpan, which utilizes a taxonomy structure and a pre-trained language model to achieve high performance. To facilitate our testing, a new fine-grained set expansion dataset is also constructed. Exper-

iments on this dataset and those used in previous studies show that FGExpan achieves significantly better performance (MAP up by 0.176) on fine-grained types and also the state-of-the-art expansion quality on coarse-grained entity sets.

Jinfeng Xiao
University of Illinois at Urbana-Champaign
jxiao13@illinois.edu

Mohab Elkaref, Nathan Herr, Geeth De Mel
IBM Research
mohab.elkaref@ibm.com, nathan.herr@ibm.com,
geeth.demel@uk.ibm.com

Jiawei Han
UIUC
hanj@illinois.edu

CP15

Mini-Batch Learning Strategies for Modeling Long Term Temporal Dependencies: A Study in Environmental Applications

Many environmental applications use recurrent neural networks (RNNs) to model physical variables with long temporal dependencies. However, due to mini-batch training, temporal relationships between training segments within the batch (intra-batch) and between batches (inter-batch) are not considered, which can lead to limited performance. Stateful RNNs consider inter-batch dependency by passing hidden states between batches but ignore intra-batch temporal dependency. To address this issue, we extend Stateful RNNs by defining a batch as a temporally ordered set of training segments, which enables intra-batch sharing of temporal information. While this strategy improves performance, its sequential training increases training times. Hence, we further propose a strategy that augments a training segment with an initial value of the target variable from the timestep right before the start of the training segment. In other words, we provide an initial value of the target variable as additional input so that the network can focus on learning changes relative to that initial value. This strategy can pass segments in any order, which reduces the training time while maintaining performance. We demonstrate the proposed strategies in hydrological modeling and observe that the most significant gains in predictive accuracy occur on the state variables whose values change more slowly, such as soil water and snowpack, rather than continuously moving flux variables such as streamflow.

Shaoming Xu, Ankush Khandelwal, Xiang Li
University of Minnesota
xu000114@umn.edu, khand035@umn.edu,
lixx5000@umn.edu

Xiaowei Jia
U. of Pittsburgh
xiaowei@pitt.edu

Licheng Liu, Jared Willard, Rahul Ghosh, Kelly Lindsay, Michael Steinbach
University of Minnesota
lichengl@umn.edu, willa099@umn.edu,
ghosh128@umn.edu, lind0436@umn.edu,
stei0062@umn.edu

Christopher Duffy
Pennsylvania State Univeristy

cxd11@psu.edu

John Nieber, Vipin Kumar
University of Minnesota
nieber@umn.edu, kumar001@umn.edu

CP15

A Linkage-Based Doubly Imbalanced Graph Learning Framework for Face Clustering

In recent years, benefiting from the expressive power of Graph Convolutional Networks (GCNs), significant breakthroughs have been made in face clustering area. However, rare attention has been paid to GCN-based clustering on imbalanced data. Although imbalance problem has been extensively studied, the impact of imbalanced data on GCN-based linkage prediction task is quite different, which would cause problems in two aspects: imbalanced linkage labels and biased graph representations. The former is similar to that in classic image classification task, but the latter is a particular problem in GCN-based clustering via linkage prediction. To tackle these challenges, we propose a linkage-based doubly imbalanced graph learning framework for face clustering. In this framework, we evaluate the feasibility of those existing methods for imbalanced image classification problem on GCNs, and present a new method to alleviate the imbalanced labels and also augment graph representations using a Reverse-Imbalance Weighted Sampling (RIWS) strategy. With the RIWS strategy, probability-based class balancing weights could ensure the overall distribution of positive and negative samples; In addition, weighted random sampling provides diverse subgraph structures, which effectively alleviates the over-fitting problem and improves the representation ability of GCNs. Extensive experiments demonstrate the effectiveness and generality of our proposed method.

Huafeng Yang, Qijie Shen, Xingjian Chen
Alibaba Group
fduspectre@gmail.com, qjshenxdu@gmail.com,
philo.cxj@alibaba-inc.com

Fangyi Zhang
Queensland University of Technology
fangyi.zhang@qut.edu.au

Rong Du
Alibaba Group
dr193339@alibaba-inc.com

CP15

Needed: Introducing Hierarchical Transformer to Eye Diseases Diagnosis

With the development of natural language processing techniques(NLP), automatic diagnosis of eye diseases using ophthalmology electronic medical records (OEMR) has become possible. It aims to evaluate the condition of both eyes of a patient respectively, and we formulate it as a particular multi-label classification task in this paper. Although there are a few related studies in other diseases, automatic diagnosis of eye diseases exhibits unique characteristics. First, descriptions of both eyes are mixed up in OEMR documents, with both free text and templated asymptomatic descriptions, resulting in sparsity and clutter of information. Second, OEMR documents contain multiple parts of descriptions and have long document lengths. Third, it is critical to provide explainability to the

disease diagnosis model. To overcome those challenges, we present an effective automatic eye disease diagnosis framework, NEEDED. In this framework, a preprocessing module is integrated to improve the density and quality of information. Then, we design a hierarchical transformer structure for learning the contextualized representations of each sentence in the OEMR document. For the diagnosis part, we propose an attention-based predictor that enables traceable diagnosis by obtaining disease-specific information. Experiments on the real dataset and comparison with several baseline models show the advantage and explainability of our framework.

Du Yi, Xu Ye
Computer Network Information Center, CAS
duyi@cnic.cn, yexu@cnic.cn

Meng Xiao
Computer Network Information Center, CAS
University of Chinese Academy of Sciences
shaow@cnic.cn

Zhiyuan Ning
Computer Network Information Center, CAS
ningzhiyuan@cnic.cn

Weiwei Dai
Changsha Aier Eye Hospital
daiweiwei@aierchina.com

Wenjuan Cui
Computer Network Information Center, CAS
wenjuancui@cnic.cn

Yuanchun Zhou
Computer Network Information Center,
Chinese Academy of Sciences
zyc@cnic.cn

CP16

Context-Aware Domain Adaptation for Time Series Anomaly Detection

Time series anomaly detection is a challenging task with a wide range of real-world applications. Due to label sparsity, training a deep anomaly detector often relies on unsupervised approaches. Recent efforts have been devoted to time series domain adaptation to leverage knowledge from similar domains. However, existing solutions may suffer from negative knowledge transfer on anomalies due to their diversity and sparsity. Motivated by the empirical study of context alignment between two domains, we aim to transfer knowledge between two domains via adaptively sampling context information for two domains. This is challenging because it requires simultaneously modeling the complex in-domain temporal dependencies and cross-domain correlations while exploiting label information from the source domain. To this end, we propose a framework that combines context sampling and anomaly detection into a joint learning procedure. We formulate context sampling into the Markov decision process and exploit deep reinforcement learning to optimize the time series domain adaptation process via context sampling and design a tailored reward function to generate domain-invariant features that better align two domains for anomaly detection. Experiments on three public datasets show promise for knowledge transfer between two similar domains and two entirely different domains.

Kwei-Herng Lai
Rice University
khlai@rice.edu

Lan Wang, Huiyuan Chen
Visa Research
lwang4@visa.com, hchen@visa.com

Kaixiong Zhou
Rice University
kaixiong.zhou@rice.edu

Fei Wang, Hao Yang
Visa Research
feiwang@visa.com, haoyang@visa.com

Xia Hu
Rice University
xia.hu@rice.edu

CP16

Matrix Profile Xxviii: Discovering Multi-Dimensional Time Series Anomalies with K of N Anomaly Detection

In recent years there has been significant progress in univariate time series anomaly detection. However, efforts to generalize this success to the multi-dimensional case have met with limited progress. The main difficulty appears to be that in any N-dimensional time series, the anomaly will generally only manifest itself on K of the time series, with $K < N$. This leads to a chicken-and-egg problem. If we knew which K time series exhibited the anomaly, it would be easy to discover its location. However, we do not know this in advance, and the search space is of size 2^N and not obviously amiable to greedy search. In this work we show a novel, simple algorithm that allows us to quickly find the best K of N anomaly subset for any value of K. Moreover, we show a simple metric that can rank the top anomaly subsets for all values of K from 1 to N. While our methods are mostly agnostic to the anomaly scoring model, for concreteness we use the Matrix Profile, and show that we can discover multi-dimensional anomalies that would escape detection by all current rival methods.

Sadaf Tafazoli
University of California Riverside
stafa002@ucr.edu

Eamonn Keogh
University of California, Riverside
eamonn@cs.ucr.edu

CP16

Deep Contrastive One-Class Time Series Anomaly Detection

The accumulation of time-series data and the absence of labels make time-series Anomaly Detection (AD) a self-supervised deep learning task. Single-normality-assumption-based methods, which reveal only a certain aspect of the whole normality, are incapable of tasks involved with a large number of anomalies. Specifically, Contrastive Learning (CL) methods distance negative pairs, many of which consist of both normal samples, thus reducing the AD performance. Existing multi-normality-

assumption-based methods are usually two-staged, firstly pre-training through certain tasks whose target may differ from AD, limiting their performance. To overcome the shortcomings, a deep Contrastive One-Class Anomaly detection method of time series (COCA) is proposed by authors, following the normality assumptions of CL and one-class classification. It treats the original and reconstructed representations as the positive pair of negative-sample-free CL, namely "sequence contrast". Next, invariance terms and variance terms compose a contrastive one-class loss function in which the loss of the assumptions is optimized by invariance terms simultaneously and the "hypersphere collapse' is prevented by variance terms. In addition, extensive experiments on two real-world time-series datasets show the superior performance of the proposed method achieves state-of-the-art.

Rui Wang
School of Computer Science and Engineering
Beihang University
ruiking@buaa.edu.cn

Chongwei Liu, Xudong Mou
Beihang University
liucw@buaa.edu.cn, mxd@buaa.edu.cn

Kai Gao
University of New South Wales
kai.gao@unsw.edu.au

Xiaohui Guo
Beihang University
guoxh@buaa.edu.cn

Pin Liu
China University of Geosciences in Beijing
liupin@cugb.edu.cn

Tianyu Wo, Xudong Liu
Beihang University
woty@act.buaa.edu.cn, liuxd@act.buaa.edu.cn

CP16
Abnormal Event Detection Via Hypergraph Contrastive Learning

Abnormal event detection, which refers to mining unusual interactions among involved entities, plays an important role in many real applications. Previous works mostly oversimplify this task as detecting abnormal pair-wise interactions. However, real-world events may contain multi-typed attributed entities and complex interactions among them, which forms an Attributed Heterogeneous Information Network (AHIN). With the boom of social networks, abnormal event detection in AHIN has become an important but seldom explored task. In this presentation, we introduce the unsupervised abnormal event detection problem in AHIN. Concretely, the events are considered star-schema instances of AHIN and are further modeled by hypergraphs. A hypergraph contrastive learning method, named AEHCL, is proposed to fully capture abnormal event patterns. AEHCL designs the intra-event and inter-event contrastive modules to exploit self-supervised AHIN information. The intra-event contrastive module captures the pair-wise and multivariate interaction anomalies within an event, and the inter-event module captures the contextual anomalies among events. During the testing phase, a contrastive learning-based abnormal event score function is further proposed to measure the abnormality degree of events. Extensive experiments on three datasets in different scenarios demonstrate the effectiveness of AEHCL.

Bo Yan, Cheng Yang, Chuan Shi, Jiawei Liu, Xiaochen Wang
Beijing University of Posts and Telecommunications
boyan@bupt.edu.cn,   yangcheng@bupt.edu.cn,
shichuan@bupt.edu.cn,   liu_jiawei@bupt.edu.cn,
2018211801@bupt.edu.cn

CP16
Subgraph Centralization: A Necessary Step for Graph Anomaly Detection

Graph anomaly detection has attracted a lot of interest recently. Despite their successes, existing detectors have at least two of the three weaknesses: (a) high computational cost which limits them to small-scale networks only; (b) existing treatment of subgraphs produces suboptimal detection accuracy; and (c) unable to provide an explanation as to why a node is anomalous, once it is identified. We identify that the root cause of these weaknesses is a lack of a proper treatment for subgraphs. A treatment called Subgraph Centralization for graph anomaly detection is proposed to address all the above weaknesses. Its importance is shown in two ways. First, we present a simple yet effective new framework called Graph-Centric Anomaly Detection (GCAD). The key advantages of GCAD over existing detectors including deep-learning detectors are: (i) better anomaly detection accuracy; (ii) linear time complexity with respect to the number of nodes; and (iii) it is a generic framework that admits an existing point anomaly detector to be used to detect node anomalies in a network. Second, we show that Subgraph Centralization can be incorporated into two existing detectors to overcome the above-mentioned weaknesses.

Zhong Zhuang
Nanjing University
Nanjing University
zhuangz@lamda.nju.edu.cn

Kai Ming Ting
Nanjing University
tingkm@lamda.nju.edu.cn

Guansong Pang
Singapore Management University
gspang@smu.edu.sg

Shuaibin Song
Nanjing University
songsb@lamda.nju.edu.cn

CP17
Estimating Propensity Scores with Deep Adaptive Variable Selection

The proliferation of observational data demands the development of statistical methods for causal inference. Many widely used causal inference methods are based on the propensity score. When estimating the propensity score, one essential question is which covariates should be included in the model. In this paper, we propose a deep adaptive variable selection based propensity score method (DAVSPS) by using representation learning and adaptive group LASSO. The key idea of DAFSPS is to combine the data-driven learning capability of representation learning and variable selection consistency of adaptive group

LASSO to improve the estimation of the propensity score by selecting confounders and adjustment variables while removing instrumental and spurious variables. We also provide a detailed theoretical analysis to prove the variable selection consistency of DAVSPS. We evaluate the performance of our method on simulated data to demonstrate its superiority over state-of-the-art methods and apply it to real data.

Zhixuan Chu
Alibaba
chuzhixuan94@gmail.com

Mechelle Claridy, Jose Cordero
University of Georgia
mechelle.claridy@uga.edu, jcordero@uga.edu

Sheng Li
University of Virginia
sheng.li@virginia.edu

Stephen Rathbun
University of Georgia
rathbun@uga.edu

CP17

Fisher Scoring Method for Neural Networks Optimization

First-order methods based on the stochastic gradient descent and variants are popularly used in training neural networks. The large dimension of the parameter space prevents the use of second-order methods in current practice. In this paper, we propose efficient ways to dynamically estimate the empirical Fisher information matrix to speed up the optimization of deep learning loss functions. We propose two different methods, both using rank-1 updates for the empirical Fisher information matrix. The first one is FisherExp and it is based on exponential smoothing using Sherman-Woodbury-Morrison matrix inversion formula. The second one is FisherFIFO, which uses a circular gradient buffer using the Sherman-Woodbury-Morrison formula twice every time a new gradient is replaced. We found that FisherFIFO scales better and we further improve scaling by proposing a partitioning strategy for the empirical Fisher Information matrix. Our methods can be used in conjunction with existing optimizers that leverage momentum-based information to improve them. We compare the performance of our methods with alternative baselines in image classification problems and found that they produce better results. Despite the overhead incurred by using second-order information, the partitioning strategy combined with parallel block updates allows us to reduce the total training time of FisherFIFO relative to the baselines.

Jackson G. De Faria
Universidade Federal de Minas Gerais
jackjufaria@gmail.com

Renato Assunção
ESRI Inc, Redlands, CA
Universidade Federal de Minas Gerais
rassuncao@esri.com

Fabricio Murai
Worcester Polytechnic Institute, Worcester, MA
fmurai@wpi.edu

CP17

A General-Purpose Transferable Predictor for Neural Architecture Search

Understanding and modelling the performance of neural architectures is key to Neural Architecture Search (NAS). Performance predictors are used in low-cost NAS and achieve high ranking correlations between predicted and ground truth performance in several NAS benchmarks. However, existing predictors are often designed based on network encodings specific to a predefined search space and are therefore not generalizable to other search spaces or new architecture families. In this paper, we propose a general-purpose neural predictor for NAS that can transfer across search spaces, by representing any given candidate Convolutional Neural Network (CNN) with a Computation Graph (CG) that consists of primitive operators. We further combine our CG network representation with Contrastive Learning (CL) and propose a graph representation learning procedure that leverages the structural information of unlabeled architectures from multiple families to train CG embeddings for our performance predictor. Experimental results on NAS-Bench-101, 201 and 301 demonstrate the efficacy of our scheme as we achieve strong positive Spearman Rank Correlation Coefficient (SRCC) on every search space, outperforming several Zero-Cost Proxies across search spaces. Moreover, when using our proposed general-purpose predictor in an evolutionary search algorithm, we can find architectures of high performance on NAS-Bench-101 and find a MobileNetV3 architecture that attains 79.2% top-1 accuracy on ImageNet.

Fred Han
Huawei Technologies Canada
fred.xuefei.han1@huawei.com

Keith G. Mills
University of Alberta
kgmills@ualberta.ca

Fabian Chudak
Huawei Technologies Canada
fchudak@gmail.com

Parsa Riahi
University of British Columbia
priahi12@gmail.com

Mohammad Salameh
Huawei Technologies Canada
mohammad.salameh@huawei.com

Jialin Zhang
Huawei Kirin Solution
zhangjialin10@hisilicon.com

Wei Lu
Huawei Technologies Canada
robin.luwei@hisilicon.com

Shangling Jui
Huawei Kirin Solution
jui.shangling@huawei.com

Di Niu
University of Alberta

dniu@ualberta.ca

CP17

Scalable Batch Acquisition for Deep Bayesian Active Learning

In deep active learning, it is especially important to choose multiple examples to markup at each step to work efficiently, especially on large datasets. At the same time, existing solutions to this problem in the Bayesian setup, such as BatchBALD, have significant limitations in selecting a large number of examples, associated with the exponential complexity of computing mutual information for joint random variables. We, therefore, present the Large Batch-BALD algorithm, which gives a well-grounded approximation to the BatchBALD method that aims to achieve comparable quality while being more computationally efficient. We provide a complexity analysis of the algorithm, showing a reduction in computation time, especially for large batches. Furthermore, we present an extensive set of experimental results on image and text data, both on toy datasets and larger ones such as CIFAR-100.

Aleksandr Rubashevskii, Daria Kotova
Skolkovo Institute of Science and Technology
aleksandr.rubashevskii@skoltech.ru, d.kotova@skoltech.ru

Maxim Panov
Technology Innovation Institute
maxim.panov@tii.ae

CP17

Multi-State Survival Analysis Using Pseudo Value-Based Deep Neural Networks

Multi-state survival analysis (MSA) uses multi-state models for analyzing time-to-event data collected from subjects who may transition to different states before experiencing the final event of interest over time. A key challenge in MSA is the accurate subject-specific prediction of multi-state model quantities, such as transition probability and state occupation probability, in the presence of censoring. The traditional statistical multi-state models typically do not use covariates, which renders them infeasible for making subject-specific predictions. Moreover, they assume a strict Markov stochastic process while modeling transition probabilities along with proportional hazard or linear covariate effect assumptions - that may not hold in real-world data. The current MSA methods have not investigated the impact of different types of censoring on the multi-state model quantities estimation. To fill the gap in the MSA literature, we propose a new class of pseudo-value-based deep learning models for MSA, where we show that pseudo values - designed to handle censoring - can be a natural replacement for estimating the subject-specific multi-state model quantities when derived from Aalen-Johansen (AJ) or Landmark AJ consistent estimators. Empirical results on both the simulated and real-world MSA datasets show that our proposed models perform better or comparably to existing MSA methods under various censoring settings and when Markovianity or linearity assumptions get violated.

Md Mahmudur Rahman
Univerversity of Maryland, Baltimore County
mrahman6@umbc.edu

Sanjay Purushotham
University of Maryland, Baltimore County
psanjay@umbc.edu

CP18

Sign-Regularized Multi-Task Learning

Multi-task learning (MTL) is a framework that enforces different tasks to share knowledge to improve generalization performance. It is a long-standing active domain that strives to handle several core issues including which tasks are correlated and similar and how to share the knowledge among correlated tasks. Existing works usually do not distinguish the polarity and magnitude of feature weights. They commonly rely on linear correlation, due to three major technical challenges 1) optimizing the models that regularize feature weight polarity, 2) deciding whether to regularize sign or magnitude, and 3) identifying which tasks should share their sign and/or magnitude patterns. To address them, this paper proposes a new multi-task learning framework that can regularize feature weight signs across tasks, beyond the conventional framework for feature weight regularization. We innovatively formulate such sign-regularization problem as a biconvex inequality-constrained optimization upon the multiplications among feature weights with slacks. We then propose a new efficient algorithm for optimization with theoretical guarantees on generalization performance and convergence. Extensive experiments on multiple datasets show the proposed methods' effectiveness, efficiency, and reasonableness of the regularized feature-weighted patterns.

Guangji Bai
Emory University
guangji.bai@emory.edu

Johnny Torres
Escuela Superior Politecnica del Litoral, ESPOL
johnnytorres83@gmail.com

Junxiang Wang, Liang Zhao
Emory University
jwan936@emory.edu, liang.zhao@emory.edu

Cristina Abad, Carmen Vaca
Escuela Superior Politecnica del Litoral, ESPOL
cabad@fiec.espol.edu.ec, cvaca@fiec.espol.edu.ec

CP18

Toward Theoretical Guidance for Two Common Questions in Practical Cross-Validation Based Hyperparameter Selection

We show, to our knowledge, the first theoretical treatments of two common questions in cross-validation based hyperparameter selection: (1) After selecting the best hyperparameter using a held-out set, we train the final model using *all* of the training data – since this may or may not improve future generalization error, should one do this? (2) During optimization such as via SGD (stochastic gradient descent), we must set the optimization tolerance ρ – since it trades off predictive accuracy with computation cost, how should one set it? Toward these problems, we introduce the *hold-in risk* (the error due to not using the whole training data), and the *model class mis-specification risk* (the error due to having chosen the wrong model class) in a theoretical view which is simple, general, and suggests heuristics that can be used when faced with a dataset instance. In proof-of-concept studies in synthetic data where theoretical quantities can be controlled, we show that these heuristics

can, respectively, (1) always perform at least as well as always performing retraining or never performing retraining, (2) either improve performance or reduce computational overhead by 2× with no loss in predictive performance.

Parikshit Ram
Georgia Institute of Technology
School of Computational Science and Engineering
p.ram@gatech.edu

Alexander Gray, Horst Samulowitz, Gregory Bramble
IBM Research
alexander.gray@ibm.com, samulowitz@us.ibm.com, bramble@us.ibm.com

CP18
Traceable Automatic Feature Transformation via Cascading Actor-Critic Agents

Feature transformation for AI is an essential task to boost the effectiveness and interpretability of machine learning (ML). Feature transformation aims to transform original data to identify an optimal feature space that enhances the performances of a downstream ML model. Existing studies either combines preprocessing, feature selection, and generation skills to empirically transform data, or automate feature transformation by machine intelligence, such as reinforcement learning. However, existing studies suffer from: 1) high-dimensional non-discriminative feature space; 2) inability to represent complex situational states; 3) inefficiency in integrating local and global feature information. To fill the research gap, we formulate the feature transformation task as an iterative, nested process of feature generation and selection, where feature generation is to generate and add new features based on original features, and feature selection is to remove redundant features to control the size of feature space. Finally, we present extensive experiments and case studies to illustrate 24.7% improvements in F1 scores compared with SOTAs and robustness in high-dimensional data.

Meng Xiao
Computer Network Information Center, CAS
University of Chinese Academy of Sciences
shaow@cnic.cn

Dongjie Wang
University of Central Florida
wangdongjie@knights.ucf.edu

Min Wu
Institute for Infocomm Research, A*STAR
wumin@i2r.a-star.edu.sg

Ziyue Qiao
The Hong Kong University of Science and Technology (Guangzho
zyqiao@ust.hk

Pengfei Wang
ICT, Chinese Academy of Science; Alibaba Group;
wpf2106@gmail.com

Kunpeng Liu
Portland State University
kunpeng@pdx.edu

Yuanchun Zhou
Computer Network Information Center,

Chinese Academy of Sciences
zyc@cnic.cn

Yanjie Fu
University of Central Florida
yanjie.fu@ucf.edu

CP18
Understanding Influence Maximization via Higher-Order Decomposition

Given its vast application on online social networks, Influence Maximization (IM) has garnered considerable attention over the last couple of decades. Due to the intricacy of IM, most current research concentrates on estimating the first-order contribution of the nodes to select a seed set, disregarding the higher-order interplay between different seeds. Consequently, the actual influence spread frequently deviates from expectations, and it remains unclear how the seed set quantitatively contributes to this deviation. To address this deficiency, this work dissects the influence exerted on individual seeds and their higher-order interactions utilizing the Sobol index, a variance-based sensitivity analysis. To adapt to IM contexts, seed selection is phrased as binary variables and split into distributions of varying orders. Based on our analysis with various Sobol indices, an IM algorithm dubbed SIM is proposed to improve the performance of current IM algorithms by over-selecting nodes followed by strategic pruning. A case study is carried out to demonstrate that the explanation of the impact effect can dependably identify the key higher-order interactions among seeds. SIM is empirically proved to be superior in effectiveness and competitive in efficiency by experiments on synthetic and real-world graphs.

Zonghan Zhang
Mississippi State University
Department of Computer Science and Engineering
zz239@msstate.edu

Zhiqian Chen
Mississippi State University
zchen@cse.msstate.edu

CP18
Learning to Learn Task Transformations for Improved Few-Shot Classification

Meta-learning has shown great promise in few-shot image classification. Many training tasks are provided to train a meta-model that can quickly learn new and similar concepts with few labeled samples. Data augmentation is often used to augment training tasks to avoid overfitting. However, existing data augmentation methods are often manually designed and fixed during training, ignoring training dynamics and the difference between various meta-learning settings specified by meta-model architectures and meta-learning algorithms. To address this problem, we add a task transformation layer between a training task and a meta-model such that the right amount of perturbation is added to training tasks for a certain meta-learning setting at a certain training stage. By jointly optimizing the task transformation layer and the meta-model, we avoid the risk of providing tasks that are either too easy or too difficult during training. We design the task transformation layer as a stochastic transformation function, adding the flexibility in how a training task can be transformed. We leverage differentiable data augmentations as the building blocks of the task transformation function for efficient optimization.

Extensive experiments show that our method can consistently improve the few-shot generalization performance of various meta-models trained with different meta-learning algorithms, meta-model architectures, and datasets.

Guangtao Zheng
University of Virginia
gz5hp@virginia.edu

Qiuling Suo
State University of New York at Buffalo
qiulings@buffalo.edu

Mengdi Huai
Iowa State University
mdhuai@iastate.edu

Aidong Zhang
University of Virginia
aidong@virginia.edu

CP19

Meta-Learning with Motif-Based Task Augmentation for Few-Shot Molecular Property Prediction

Predicting molecular properties is known to be a difficult few-shot learning problem in drug discovery due to high failure rates of virtual screening. Meta-learning frameworks have been successfully leveraged to tackle this few-shot molecular property prediction problem. Recent work has shown that the generalization of meta-learning can be further improved by task augmentation techniques. However, it is challenging to directly apply current popular task augmentation strategies to molecular modeling due to its highly discrete nature. To this end, we propose a motif-based task augmentation (MTA) technique to address this challenge. MTA is mainly conducting task augmentations by generating new labeled samples through retrieving highly relevant motifs from a pre-defined motif vocabulary. Both support samples and query samples are interpolated with their corresponding retrieved motifs in the latent space in a convex manner parameterized by a specific probability distribution. Augmented query samples are classified by measuring their Euclidean distances to augmented prototype clusters. With our novel MTA approach, the above two overfitting issues can be alleviated through establishing connections between different tasks with augmented motifs. Empirical results for popular benchmark datasets demonstrate that our approach can consistently outperform existing baseline methods and can also provide some interpretable relation information between structural motifs and molecules.

Ziqiao Meng, Yaoman Li
The Chinese University of Hong Kong
ziqiao.meng@mail.utoronto.ca, ymli@cse.cuhk.edu.hk

Peilin Zhao, Yang Yu
Tencent AI Lab
masonzhao@tencent.com, kevinyyu@tencent.com

Irwin King
The Chinese University of Hong Kong
king@cse.cuhk.edu.hk

CP19

Reinforcement Learning Guided Multi-Objective

Exam Paper Generation

To reduce the repetitive and complex work of instructors, exam paper generation (EPG) technique has become a salient topic in the intelligent education field, which targets at generating high-quality exam paper automatically according to instructor-specified assessment criteria. The current advances utilize the ability of heuristic algorithms to optimize several well-known objective constraints, such as difficulty degree, number of questions, etc., for producing optimal solutions. However, in real scenarios, considering other equally relevant objectives (e.g., distribution of exam scores, skill coverage) is extremely important. Besides, how to develop an automatic multi-objective solution that finds an optimal subset of questions from a huge search space of large-sized question datasets and thus composes a high-quality exam paper is urgent but nontrivial. To this end, we skillfully design a reinforcement learning guided Multi-Objective Exam Paper Generation framework, termed MOEPG, to simultaneously optimize three exam domain-specific objectives including difficulty degree, distribution of exam scores, and skill coverage. Specifically, to accurately measure the skill proficiency of the examinee group, we first employ deep knowledge tracing to model the interaction information between examinees and response logs. We then design the flexible Exam Q-Network, a function approximator, which automatically selects the appropriate question to update the exam paper composition process. Later, MOEPG divides the decision space into multiple subspaces to better guide the updated direction of the exam paper. Through extensive experiments on two real-world datasets, we demonstrate that MOEPG is feasible in addressing the multiple dilemmas of exam paper generation scenario.

Yuhu Shang
Tianjin University of Science & Technology, China
shangyuhu6902@mail.tust.edu.cn

Xuexiong Luo
Macquarie University
xuexiong.luo@hdr.mq.edu.au

Lihong Wang
National Computer Network Emergency Response
Technical Team
wlh@isc.org.cn

Hao Peng
Beihang University, China
penghao@buaa.edu.cn

Xiankun Zhang
Tianjin University of Science and Tech
zhxkun@tust.edu.cn

Yimeng Ren
Tianjin University of Science & Technology, China
renyimeng@mail.tust.edu.cn

Kun Liang
liangkun@tust.edu.cn
liangkun@tust.edu.cn

CP19

Probabilistic Inverse Modeling: An Application in Hydrology

Rapid advancement in inverse modeling methods have

brought into light their susceptibility to imperfect data. This has made it imperative to obtain more explainable and trustworthy estimates from these models. In hydrology, basin characteristics can be noisy or missing, impacting streamflow prediction. We propose a probabilistic inverse model framework that can reconstruct robust hydrology basin characteristics from dynamic input weather driver and streamflow response data. We address two aspects of building more explainable inverse models, uncertainty estimation (uncertainty due to imperfect data and imperfect model) and robustness. This can help improve the trust of water managers, handling of noisy data and reduce costs. We also propose an uncertainty-based loss regularization that offers removal of 17% of temporal artifacts in reconstructions, 36% reduction in uncertainty and 4% higher coverage rate for basin characteristics. The forward model performance (streamflow estimation) is also improved by 6% using this uncertainty learning based reconstructions.

Somya Sharma, Rahul Ghosh, Arvind Renganathan, Xiang Li
University of Minnesota
sharm636@umn.edu, ghosh128@umn.edu, renga016@umn.edu, lixx5000@umn.edu

Snigdhansu Chatterjee
University of Minnesota
Department of Statistics
chatt019@umn.edu

John Nieber
University of Minnesota
nieber@umn.edu

Christopher Duffy
Pennsylvania State Univeristy
cxd11@psu.edu

Vipin Kumar
University of Minnesota
kumar001@umn.edu

CP19

A Novel Reject Option Applied to Sleep Stage Scoring

Sleep stage scoring is an essential component of diagnosing sleep disorders. Unfortunately, it is a time-intensive task that requires clinical experts to annotate an entire nights recording for each patient. Therefore, machine learned models offer the potential to alleviate this burden by automating this task. While learned models achieve acceptable accuracy on curated data, these models still produce highly inaccurate scorings for certain patients when deployed in medical centers. This is because particular subsets of the population may not be adequately represented in the data used to train the model. For example, data are not easily accessible or are hard or impossible to collect. This creates trust issues as incorrect scorings can have severe consequences such as undetected diseases. To address this, we propose augmenting an existing model with a reject option which enables it to abstain from making predictions if the model is at an elevated risk of making a mistake. We show that traditional rejection frameworks can systematically be too cautious in certain circumstances and abstain even when the model can make good predictions. We propose a solution by considering both the data distribution and the model predictions. We demonstrate the efficacy of our method on a real-world sleep scoring use case. Moreover, we found that our approach leads to improved performance on several publicly available benchmarks

Dries Van der Plas, Wannes Meert
KU Leuven
dries.vanderplas@kuleuven.be, wannes.meert@kuleuven.be

Johan Verbraecken
University of Antwerp
Antwerp University Hospital
johan.verbraecken@uza.be

Jesse Davis
KU Leuven
jesse.davis@kuleuven.be

CP19

Eco-Pinn: A Physics-Informed Neural Network for Eco-Toll Estimation

The eco-toll estimation problem quantifies the expected environmental cost (e.g., energy consumption, exhaust emissions) for a vehicle to travel along a path. This problem is important for societal applications such as eco-routing, which aims to find paths with the lowest exhaust emissions or energy need. The challenges of this problem are three-fold: (1) the dependence of a vehicle's eco-toll on its physical parameters; (2) the lack of access to data with eco-toll information; and (3) the influence of contextual information (i.e. the connections of adjacent segments in the path) on the eco-toll of road segments. Prior work on eco-toll estimation has mostly relied on pure data-driven approaches and has high estimation errors given the limited training data. To address these limitations, we propose a novel Eco-toll estimation Physics-informed Neural Network framework (Eco-PiNN) using three novel ideas, namely, (1) a physics-informed decoder that integrates the physical laws governing vehicle dynamics into the network, (2) an attention-based contextual information encoder, and (3) a physics-informed regularization to reduce overfitting. Experiments on real-world heavy-duty truck data show that the proposed method can greatly improve the accuracy of eco-toll estimation compared with state-of-the-art methods.

Yan Li, Mingzhou Yang, Matthew Eagon, Majid Farhadloo
University of Minnesota, Twin Cities
lixx4266@umn.edu, yang7492@umn.edu, eagon012@umn.edu, farha043@umn.edu

Yiqun Xie
University of Maryland
xie@umd.edu

William F. Northrop
University of Minnesota, Twin Cities
wnorthro@umn.edu

Shashi Shekhar
University of Minnesota
shekhar@umn.edu

CP20

TEM: High Utility Metric Differential Privacy on

Text

Ensuring the privacy of users whose data are used to train Natural Language Processing (NLP) models is necessary to build and maintain customer trust. Differential Privacy (DP) has emerged as the most successful method to protect the privacy of individuals. However, applying DP to the NLP domain comes with unique challenges. The most successful previous methods use a generalization of DP for metric spaces, and apply the privatization by adding noise to inputs in the metric space of word embeddings. However, these methods assume that one specific distance measure is being used, ignore the density of the space around the input, and assume the embeddings used have been trained on public data. In this work we propose Truncated Exponential Mechanism (TEM), a general method that allows the privatization of words using any distance metric, on embeddings that can be trained on sensitive data. Our method makes use of the exponential mechanism to turn the privatization step into a selection problem. This allows the noise applied to be calibrated to the density of the embedding space around the input, and makes domain adaptation possible for the embeddings. In our experiments, we demonstrate that our method outperforms the state-of-the-art in terms of utility for the same level of privacy, while providing more flexibility in the metric selection.

Ricardo Silva Carvalho
Simon Fraser University
rsilvaca@sfu.ca

Theodore Vasiloudis, Oluwaseyi Feyisetan
Amazon
theodoros.vasiloudis@gmail.com, sey@amazon.com

Ke Wang
Simon Fraser University
wangk@sfu.ca

CP20

Group AdaBoost with Fairness Constraint

Boosting is a widely-used ensemble learning technique that combines a series of weak base learners sequentially. While it can achieve competitive predictive performance, boosting may induce unfairness within the ensemble process when applied to group-unbalanced data, which can pose a major obstacle for using boosting in human-centered predictions such as in health informatics and risk management. To address this issue, we propose a boosting-based framework that includes rectified fairness penalty, penalty intensity selection, and post-pruning. This framework can provide a certified fairness guarantee by stabilizing the ensemble process with a fairness constraint given by users, resulting in not only improved fairness but also higher averaged accuracy among different groups. In five widely-used fairness benchmarks, we demonstrate the effectiveness of our proposed method against various baselines aiming for group fairness. On average, our method outperforms the best baseline of fairness metrics such as equalized odds, with higher accuracy at the same time. Also, by converting the baselines into different group imbalanced degrees, our method is demonstrated to be better and more stable comparing existing baselines. The code is available at https://github.com/chrisyxue/GAFairChere.

Zhiyu Xue
CSE
Michigan State University
xuezhiyu@msu.edu

CP20

Pmp: Privacy-Aware Matrix Profile Against Sensitive Pattern Inference for Time Series

Recent rapid development of sensor technology has allowed massive fine-grained time series (TS) data to be collected and set the foundation for the development of data-driven services. During the process, data sharing is often involved to allow the third-party modelers to perform specific time series data mining tasks. The high resolution of TS brings new challenges in protecting privacy. Numerous research have found that long shape-based patterns could contain more sensitive information and may potentially be extracted and misused by a malicious third party. However, the privacy issue for TS patterns is surprisingly seldom explored. In this work, we consider a new privacy-preserving problem: preventing malicious inference on long shape-based patterns while preserving short segment information for the utility task. To mitigate the challenge, we investigate an alternative approach by sharing Matrix Profile (MP), which is a non-linear transformation of original data and a versatile data structure that supports many data mining tasks. We found that while MP can prevent concrete shape leakage, the canonical correlation in MP index can still reveal the location of sensitive long pattern. To protect MP from attacks, we propose a Privacy-Aware Matrix Profile (PMP) via perturbing the local correlation and breaking the canonical correlation in MP index vector. We evaluate PMP through quantitative analysis and real-world case studies to show the effectiveness of the proposed method.

Li Zhang
George Mason University
lzhang18@gmu.edu

Jiahao Ding
University of Houston
jding7@uh.edu

Yifeng Gao
University of Texas Rio Grande Valley
yifeng.gao@utrgv.edu

Jessica Lin
George Mason University
jessica@gmu.edu

CP20

Fairness-Aware Multi-View Clustering

In the era of big data, we are often facing the challenge of data heterogeneity and the lack of label information simultaneously. In the financial domain (e.g., fraud detection), the heterogeneous data may include not only numerical data (e.g., total debt), but also text and images (e.g., financial statement). At the same time, the label information (e.g., fraud transactions) may be missing for building predictive models. To address these challenges, many state-of-the-art multi-view clustering methods have been proposed and achieved outstanding performance. However, these methods typically do not take into consideration the fairness aspect and are likely to generate biased results using sensitive information such as race and gender. Therefore, in this paper, we propose a fairness-aware multi-view clustering method named FairMVC. It incorpo-

rates the group fairness constraint into the soft membership assignment for each cluster to ensure that the fraction of different groups in each cluster is approximately identical to the entire data set. Meanwhile, we adopt the idea of both contrastive learning and non-contrastive learning and propose novel regularizers to handle heterogeneous data in complex scenarios with missing data or noisy features. Experimental results on real-world data sets demonstrate the effectiveness and efficiency of the proposed framework. We also derive insights regarding the relative performance of the proposed regularizers in various scenarios.

Lecheng Zheng
University of Illinois at Urbana-Champaign
lecheng4@illinois.edu

Yada Zhu
IBM T.J. Watson Research Center
yzhu@us.ibm.com

Jingrui He
University of Illinois at Urbana-Champaign, U.S.
jingrui@illinois.edu

CP20
On Improving Fairness of AI Models with Synthetic Minority Oversampling Techniques

Biased AI models result in unfair decisions. In response, a number of algorithmic solutions have been engineered to mitigate bias, among which the Synthetic Minority Oversampling Technique (SMOTE) has been studied, to an extent. Although the SMOTE technique and its variants have great potentials to help improve fairness, there is little theoretical justification for its success. In addition, formal error and fairness bounds are not clearly given. This paper attempts to address both issues. We prove and demonstrate that synthetic data generated by oversampling underrepresented groups can mitigate algorithmic bias in AI models, while keeping the predictive errors bounded. We further compare this technique to the existing state-of-the-art fair AI techniques on five datasets using a variety of fairness metrics. We show that this approach can effectively improve fairness even when there is a significant amount of label and selection bias, regardless of the baseline AI algorithm.

Yan Zhou, Murat Kantarcioglu
University of Texas at Dallas
yan.zhou2@utdallas.edu, muratk@utdallas.edu

Chris Clifton
Department of Computer Science
Purdue University
clifton@cs.purdue.edu

CP21
Gotta: Generative Few-Shot Question Answering by Prompt-Based Cloze Data Augmentation

Few-shot question answering (QA) aims at precisely discovering answers to a set of questions from context passages while only a few training samples are available. Although existing studies have made some progress and can usually achieve proper results, they suffer from understanding deep semantics for reasoning out the questions. In this paper, we develop GOTTA, a **G**enerative pr**O**mp**T**-based da**T**a **A**ugmentation framework to mitigate the challenge above. Inspired by the human reasoning process, we propose to integrate the cloze task to enhance few-shot QA learning. Following the recent success of prompt-tuning, we present the cloze task in the same format as the main QA task, allowing the model to learn both tasks seamlessly together to fully take advantage of the power of prompt-tuning. Extensive experiments on widely used benchmarks demonstrate that GOTTA consistently outperforms competitive baselines, validating the effectiveness of our proposed prompt-tuning-based cloze task, which not only fine-tunes language models but also learns to guide reasoning in QA tasks. Further analysis shows that the prompt-based loss incorporates the auxiliary task better than the multi-task loss, highlighting the strength of prompt-tuning on the few-shot QA task. In this presentation, I will introduce in detail the technical contributions of this paper.

Xiusi Chen
University of California, Los Angeles
xchen@cs.ucla.edu

Yu Zhang
University of Illinois at Urbana-Champaign
yuz9@illinois.edu

Jinliang Deng
University of Technology Sydney
jinliang.deng@student.uts.edu.au

Jyun-Yu Jiang
Amazon Search
jyunyu@amazon.com

Wei Wang
University of California Los Angeles
weiwang@cs.ucla.edu

CP21
Coarse-to-Fine Open Information Extraction via Relation Oriented Reading Comprehension

Open information extraction (Open IE), aiming at distilling structured, machine-readable triples from natural language text, plays an important role in various applications, including natural language understanding, knowledge graph construction, etc. Previous supervised Open IE approaches are mostly tailored to extract predicate-argument triples, in which the predicate is usually limited to verb phrases, whereas, the semantic relations expressed within noun phrases are being neglected. However, identifying semantic relation between entities is no trivial task due to the implicit and complex relation expressions. To address the above issue, we present ReadOIE, a framework for coarse-to-fine Open IE via relation oriented reading comprehension, for extracting relation-entity triples. In our framework, all entity pairs are extracted to generate structured questions and the input sentence is regarded as the context passage. Semantic relations that best answer the questions are then extracted by comprehending the given context. Moreover, in order to identify the non-existence relations between entities, we design a coarse-to-fine relation extraction approach consisted of an extensive detection module and an intensive extraction module. The extensive detection performs relation existence judgement on a coarse level and intensive extraction identifies the relation on a fine-grained level. Extensive experiments on benchmark datasets show that ReadOIE outperforms the SOTA baselines.

Rui Meng

Guangdong Provincial Key Laboratory of IRADS
BNU-HKBU United International College
ruimeng@uic.edu.cn

Tingxin Li
Cornell University
tl687@cornell.edu

Feng Chen
The University of Adelaide
chenfeng1271@gmail.com

Jianming Wu
Guangzhou University
jianmingwu@e.gzhu.edu.cn

CP21
XDC: Adaptive Cross Domain Short Text Clustering

Short text clustering is a challenging unsupervised learning task which requires a complex representation of each document to effectively model the semantics and syntactic structure of the text. Existing works have attempted to tackle this challenging task by incorporating additional information to the model, such as number of clusters, number of datapoints in each clusters, the distribution of the input data, and more. Unlike previous approaches, we propose to exploit an auxiliary dataset that is fully labeled to augment the quality of the learned representations. We also define the problem as cross domain clustering (XDC), which leverages adversarial learning to train an adaptive clustering model across text domains. Specifically, XDC jointly exploits a labeled source domain and an unlabeled target domain during model training. Owing to domain adversarial learning, the distribution shift across source and target domains could be mitigated. Moreover, XDC is implemented as a linkage-based clustering approach using graphs, which is agnostic of the number of clusters. We evaluate our XDC framework on three text datasets, and results show that it outperforms the state-of-the-art text clustering methods in most cases. Ablation studies and qualitative analysis also demonstrate the effectiveness of our framework.

Saed Rezayi
University of Georgia
saedr@uga.edu

Handong Zhao
Adobe
hazhao@adobe.com

Ronghang Zhu
University of Georgia
ronghangzhu@uga.edu

Sheng Li
University of Virginia
shengli@virginia.edu

CP21
Hierarchical Neural Topic Model with Embedding Cluster and Neural Variational Inference

Compared to flat topic models, hierarchical topic models not only exploit inherent structural information in the corpus but detect better semantic topics with the help of hierarchy knowledge. Recently, Neural-Variational-Inference (NVI) based hierarchical neural topic models have achieved better performance. However, existing NVI-based models learn topics of different levels with the same strategy, i.e., word co-occurrence patterns, which causes that topics of different levels cannot be distinguished from a semantic perspective and topics of the first level degenerate into some meaningless common words. To address the above problems, we propose a novel Hierarchical Neural Topic Model with embedding cluster and neural variational inference (C-HNTM). Specifically, C-HNTM adopts Gaussian Mixture Model (GMM) to learn topics of the first level based on word embeddings, which can capture the global semantic information of the whole corpus and generate more meaningful and global semantic topics. Then, the NVI-based method is adopted to learn topics of the second level with Bag-of-Word from a document perspective, which can generate local and more detailed topics. Third, we simultaneously learn global and local topic distributions and dependency matrix by using Stochastic Gradient Variational Bayes (SGVB) estimator. Finally, we provide the detailed inference of variational lower bound and extensive experiments on three real-world datasets to validate the effectiveness of our model.

Ningjing Wang
Beihang University
wn1j@qq.com

Deqing Wang
Beihang University
Zhongguancun Laboratory
dqwang@buaa.edu.cn

Ting Jiang, Chenguang Du, Chuyu Fang
Beihang University
royokong@buaa.edu.cn, duchenguang@buaa.edu.cn,
fangcybh@buaa.edu.cn

Fuzhen Zhuang
Beihang University
Zhongguancun Laboratory
zhuangfuzhen@buaa.edu.cn

CP21
Feature Enhanced Zero-Shot Stance Detection via Contrastive Learning

Zero-shot stance detection is challenging because it requires detecting the stance of previously unseen targets in the inference phase. The ability to learn transferable target-invariant features is critical for zero-shot stance detection. In this paper, we propose a stance detection approach that can efficiently adapt to unseen targets, the core of which is to capture target-invariant syntactic expression patterns as transferable knowledge. Specifically, we first augment the data by masking the topic words of sentences, and then feed the augmented data to an unsupervised contrastive learning module to capture transferable features. Besides, to fit a specific target, we encode the raw text as target-specific features. Finally, we adopt an attention mechanism, which combines syntactic expression patterns with target-specific features to obtain enhanced features for predicting previously unseen targets. Experiments demonstrate that our model outperforms competitive baselines on four benchmark datasets.

Xuechen Zhao, Jiaying Zou, Zhong Zhang, Feng Xie, Bin Zhou, Lei Tian

National University of Defense Technology
zhaoxuechen@nudt.edu.cn, zoujiaying20@nudt.edu.cn, zhangzhong@nudt.edu.cn, xiefeng@nudt.edu.cn, binzhou@nudt.edu.cn, leitian129@nudt.edu.cn

CP22
Towards Trustworthy Representation Learning

Representation learning (RL) aims to extract latent features from various types of data and then facilitate a wide range of downstream data analytics tasks, such as classification, clustering, outlier detection, recommender systems, etc. Prior efforts on RL in the past decades mainly focus on developing models to largely retain useful information (e.g., discriminative patterns, semantic knowledge) from data while discard redundant and noisy information. In the data mining and machine learning communities, some recent efforts attempt to encourage both task-oriented performance and trustworthiness of the model, suggesting a pathway towards trustworthy representation learning (TRL). Although trustworthiness has been increasingly discussed from different perspectives (e.g., fairness, explainability, robustness), the intertwined connections between representation learning and trustworthiness have not been formally discussed and clearly revealed yet. In this Blue Sky vision paper, for the first time, we present a conceptual framework that illustrates how to characterize trustworthiness in representation learning, discuss the research challenges, and point out future research opportunities.

Sheng Li
University of Virginia
vga8uf@virginia.edu

CP22
Statistically-Sound Knowledge Discovery from Data

Knowledge Discovery from Data (KDD) has mostly focused on understanding the available data. *Statistically-sound* KDD shifts the goal to understanding the partially unknown, random Data Generating Process (DGP) process that generates the data. This shift is necessary to ensure that the results from data analysis constitute new knowledge about the DGP, as required by the practice of scientific research and by many industrial applications, to avoid costly *false discoveries*. In statistically-sound KDD, results obtained from the data are considered as *hypotheses*, and they must undergo *statistical testing*, before being deemed significant, i.e., informative about the DGP. The challenges include (1) how to subject the hypotheses to *severe testing* to make it hard for them to be deemed significant; (2) considering the simultaneous testing of *multiple hypotheses* as the default setting, not as an afterthought; (3) offering *flexible statistical guarantees* at different stages of the discovery process; and (4) achieving *scalability along multiple axes*, from the size of the data to the number and complexity of hypotheses to be tested. Success for Statistically-sound KDD as a field will be achieved with (1) the introduction of a rich collection of null models that are representative of the KDD tasks, and of the existing knowledge of the GDP by field experts; (2) the development of scalable algorithms for testing results for many KDD tasks on different data types; and (3) the availability of benchmark dataset generators that allow to thoroughly evaluate these algorithms.

Matteo Riondato
Amherst College

mriondato@amherst.edu

CP22
Missed Opportunities in Fair AI

Nripsuta Saxena
USC
nsaxena@usc.edu

CP22
Making a Computational Attorney

This "blue sky idea" paper outlines the opportunities and challenges in data mining and machine learning involving making a computational attorney — an intelligent software agent capable of helping human lawyers with a wide range of complex high-level legal tasks such as drafting legal briefs for the prosecution or defense in court. In particular, we discuss what a ChatGPT-like Large Legal Language Model (L3M) can and cannot do today, which will inspire researchers with promising short-term and long-term research objectives.

Frank Schilder
Thompson Reuters Lab
frank.schilder@thomsonreuters.com

Dell Zhang
Thomson Reuters Labs
dell.z@ieee.org

Jack G. Conrad, Masoud Makrehchi, David von Rickenbach, Isabelle Moulinier
Thompson Reuters Lab
jackgconrad@gmail.com, masoud.makrehchi@thomsonreuters.com, david.vonrickenbach@thomsonreuters.com, isabelle.moulinier@thomsonreuters.com

CP22
Data Mining Challenges and Opportunities to Achieve Net Zero Carbon Emissions: Focus on Electrified Vehicles

Society must achieve net zero carbon emissions to mitigate anthropogenic climate change and preserve a livable planet. Reducing transportation emissions is an important component to achieve net zero because such emissions account for a quarter of global carbon released into the environment. Driven by increasingly available transportation big data and enhanced computational speed, data mining techniques have become powerful tools to achieve transportation decarbonization. This paper describes existing gaps in transportation decarbonization research where data mining can help address problems related to medium and heavy vehicle electrification, electric micromobility safety, and analysis of alternative fuel-powered and plug-in hybrid electric vehicles. Our recommendations encompass open research problems, opportunities for data mining applications, and examples of areas where advancements in data mining techniques are needed. We encourage the data mining community to explore these challenges and opportunities to help achieve net zero emissions goals.

Mingzhou Yang, Bharat Jayaprakash, Matthew Eagon, Hyeonjung (Tari) Jung, William F. Northrop
University of Minnesota, Twin Cities

yang7492@umn.edu, jaya015@umn.edu,
eagon012@umn.edu, jungx367@umn.edu,
wnorthro@umn.edu

Shashi Shekhar
University of Minnesota, Twin-Cities
shekhar@cs.umn.edu

CP22

Data-Centric AI: Perspectives and Challenges

The role of data in building AI systems has recently been significantly magnified by the emerging concept of data-centric AI (DCAI), which advocates a fundamental shift from model advancements to ensuring data quality and reliability. Although our community has continuously invested efforts into enhancing data in different aspects, they are often isolated initiatives on specific tasks. To facilitate the collective initiative in our community and push forward DCAI, we draw a big picture and bring together three general missions: training data development, inference data development, and data maintenance. We provide a top-level discussion on representative DCAI tasks and share perspectives. Finally, we list open challenges to motivate future exploration. A companion list of data-centric AI resources will be regularly updated on https://github.com/daochenzha/data-centric-AI

Daochen Zha
Rice University
daochen.zha@rice.edu

Zaid Pervaiz Bhat
Texas A&M University
zaid.bhat1234@tamu.edu

Kwei-Herng Lai
Rice University
khlai@rice.edu

Fan Yang
Rice
fyang@rice.edu

Xia Hu
Rice University
xia.hu@rice.edu

MT1

See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT2

See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.

N/A

MT3
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT4
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT5
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT6
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT7
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT8
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A

MT9
See Tutorial Webpage for Tutorial Information

Tutorial Attendees
.
N/A